HONORS ABSTRACT ALGEBRA AT PEKING UNIVERSITY

FALL 2023

This is an ongoing project to develop a series of lecture notes for the Honors Abstract Algebra course at Peking University.

0. Basic structures in algebra

The purpose of this section is to collect and review basic definitions of some algebraic concepts that we assume the students have seen prior to the course, for example, fields, vector spaces, and etc. These will be reintroduced throughout the course, but before that, we might need to refer to them from time to time for the purpose of examples.

Notation 0.0.1. The following notations will be used without definition:

- $\mathbb{Z}$ denotes the set/group/ring of integers.
- $\mathbb{Q}$ denotes the set/field of rational numbers.
- $\mathbb{R}$ denotes the set/field of real numbers.
- $\mathbb{C}$ denotes the set/field of complex numbers.
- For a set S, $|S|$ denotes its cardinality.

One big difference in notation from calculus is that we often use g' to denote the derivative of g in calculus. We NEVER do that in abstract algebra. Often g' denotes just another element, often one that shares similar properties as g.

0.1. **Binary operations and groups.**

Definition 0.1.1. Let S be a set, a **binary operation** on S is a map $* : S \times S \to S$. For $a, b \in S$, instead of writing $*(a, b)$, we write $a * b$ instead.

Example 0.1.2. Let S denote all continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Then we have the following binary operations: for $f, g \in S$,

- <u>addition</u>: $f + g$ given by $(f + g)(x) = f(x) + g(x)$ for $x \in \mathbb{R}$;
- <u>subtraction</u>: $f - g$ given by $(f - g)(x) = f(x) - g(x)$ for $x \in \mathbb{R}$;
- <u>multiplication</u>: $f \cdot g$ given by $(f \cdot g)(x) = f(x) \cdot g(x)$ for $x \in \mathbb{R}$;
- <u>composition</u>: $f \circ g$ given by $(f \circ g)(x) = f(g(x))$ for $x \in \mathbb{R}$.

Definition 0.1.3. A **group** is a pair of a nonempty set G and a binary operation $*$ on G such that

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

The group G is called **abelian** or **commutative** if $a * b = b * a$ for all $a, b \in G$.

Remark 0.1.4. When G is commutative, we often write $a + b$ for the binary operation.

Definition 0.1.5. A **subgroup** of a group $(G, *)$ is a subset H such that for every $a, b \in H$

$$ab^{-1} \in H.$$

See later in Section 1 for a more detailed introduction of the concept of groups and subgroups.

0.2. Rings and fields.

Definition 0.2.1. A **ring** R is a nonempty set together with two binary operations $+$ and $\cdot$, satisfying

(1) $(R, +)$ is an abelian group (with additive identity denoted by 0);
(2) the operator $\cdot$ is *associative*, i.e. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for $a, b, c \in R$;
(3) the **distributive law** holds in R, i.e. for all $a, b, c \in R$,

$$(a + b) \cdot c = a \cdot c + b \cdot c, \quad \text{and} \quad a \cdot (b + c) = a \cdot b + a \cdot c;$$

(4) there exists an element $1 \in R$ such that $0 \neq 1$, called **unity**, such that for any $a \in R$,

$$1 \cdot a = a = a \cdot 1.$$

In particular, throughout this course, all rings are assumed to have 1 and we will always assume that $0 \neq 1$. This is slightly different from some other references.

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

Notation 0.2.2. Let R be a ring. An element $r \in R$ is called a **unit** if there exists $s \in R$ such that $r \cdot s = s \cdot r = 1$. The set of all units in R, denoted by $R^\times$, is a group under the multiplication with 1 as the identity element.

The following example will be frequently used later.

Example 0.2.3. Fix a positive integer n. Let $\mathbf{Z}_n = \{\bar{0}, \bar{1}, \dots, \overline{n-1}\}$ denote the full set of residual classes modulo n, where $\bar{i}$ is short for i mod n. In general, for $i \in \mathbb{N}$, we write $\bar{i} = \bar{i}_0$ for the unique number $i_0 \in \{0, \dots, n-1\}$ such that $i \equiv i_0$ mod n. One can add or multiply two residual classes, writing $+_n$ for this addition and $\times_n$ for this multiplication, namely

$$\bar{a} +_n \bar{b} = \overline{a + b}, \qquad \bar{a} \times_n \bar{b} = \overline{ab}.$$

This defines an commutative ring $(\mathbf{Z}_n, +_n, \times_n)$.

The set of units in $\mathbf{Z}_n$ is $\mathbf{Z}_n^\times = \{\bar{i} \mid \gcd(i, n) = 1\}$. Its order, denoted by $\varphi(n)$, is called the *Euler function.*

Note: we will try to distinguish $\mathbf{Z}_n$ from $\mathbb{Z}_n$; this is because writing $\mathbb{Z}_p$ with p a prime number often means p-adic numbers in number theory.

Definition 0.2.4. A commutative ring is called a **field** if every nonzero element $a \in R$ has a multiplicative inverse.

Example 0.2.5. The set of integers $\mathbb{Z}$ with the natural addition and multiplication is a commutative ring.

The set of rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, and complex numbers $\mathbb{C}$ are fields and are in particular commutative rings.

Example 0.2.6. The ring $\mathbf{Z}_n$ introduced in Example 0.2.3 is in fact a field when $n = p$ is a prime number. In this case, we write $\mathbb{F}_p$ for it instead (to emphasize that this is a field). We will prove later in this book that for each power of a prime p^n, there exists exactly one finite field, which we denote by $\mathbb{F}_{p^n}$. Finite fields provide a large source of examples in abstract algebra.

0.3. **Modules.**

Definition 0.3.1. Let R be a ring. A **left R-module** is an abelian group M with a map (called the left R-action)

$$\begin{aligned} R \times M &\longrightarrow M \\ (a, m) &\longmapsto a \cdot m, \end{aligned} \tag{0.3.1.1}$$

satisfying the following conditions: for $r, s \in R$ and $m, n \in M$,

(1) $1 \cdot m = m$;
(2) $(r + s) \cdot m = r \cdot m + s \cdot m$;
(3) $r \cdot (m + n) = r \cdot m + r \cdot n$;
(4) $r \cdot (s \cdot m) = (r \cdot s) \cdot m$.

When R is a field, a left R-module M is called a **vector space**.

Example 0.3.2. Let R be a ring, R maybe viewed as a left R-module when the left R-action (0.3.1.1) is given by left multiplication.

More generally, for $r \in \mathbb{N}$, let M denote $\{(m_1, \dots, m_r) \,|\, m_1, \dots, m_r \in R\}$, the set of r-tuples in R, equipped with an abelian group structure given by termwise addition, namely

$$(m_1, \dots, m_r) + (m'_1, \dots, m'_r) = (m_1 + m'_1, \dots, m_r + m'_r).$$

Then M admits a left R-module structure given by

$$a \cdot (m_1, \dots, m_r) = (a \cdot m_1, \dots, a \cdot m_r).$$

1. Groups and subgroups

Before we start the journey to the abstract world, we hope the students keep in mind of one thing: before diving into an abstract mathematical concept, we should ask: why should we care about such a concept? What are important examples, typical examples, and also counterexamples? Making things abstract for the purpose of making them abstract is not mathematics.

1.1. **Why do we study groups?** Loosely speaking, the concept of groups originated in:

- describing "symmetry", *uniformly* for different situations, and
- *comparing* symmetries in different context, abstractly.

We only give two examples here, and we will see more later in this course.

Example 1.1.1. In the first example, we discuss the five platonic solids:

Tetrahedron, Hexahedron, Octahedron, Dodecahedron, and Isocahedron.

The symmetry of each platonic solid is the set of ways to rotate (and possibly reflect) the space at the center of the solid yet keeping the solid stable. For example, if we label the four vertices of the tetrahedron as 1, 2, 3, and 4, then each such symmetry will induce a way to permute these numbers $\{1, 2, 3, 4\}$, establishing relations between symmetries of the solids and permutations of its vertices.

By giving our three dimensional space a Cartesian coordinate system centered at the center of a platonic solid, we may represent each symmetry as a 3×3 real matrix, and all symmetries form a *finite* group in $\mathrm{GL}_3(\mathbb{R})$. In some sense, the classification of platonic solids comes hand-in-hand with the classification of finite subgroups of $\mathrm{GL}_3(\mathbb{R})$.

Going towards higher dimensional situations when our three-dimensional intuition presents limitations, abstract group theory will play a much stronger role.

Example 1.1.2. The second example comes from Diophantine equations. Let $D > 1$ be a square free integer. It is a classical number theory theorem that the set of integer solutions to the *Pell's equation*:

$$x^2 - Dy^2 = 1 \tag{1.1.2.1}$$

can be described as follows: there exists a "fundamental solution" $(x_0, y_0) \in \mathbb{Z}^2$, and then all other solutions to (1.1.2.1) can be deduced by writing for $n \in \mathbb{Z}$,

$$(x_0 + \sqrt{D}y_0)^n = x_n + \sqrt{D}y_n, \qquad \text{with } x_n, y_n \in \mathbb{Z}.$$

Then $\{\pm(x_n, y_n) \in \mathbb{Z}^2 \,|\, n \in \mathbb{Z}\}$ are all the integer solutions to (1.1.2.1). In other words, the integer solutions to the Pell's equation admit some addition structure that "looks like" $\{\pm\} \times \mathbb{Z}$.

If we summarize the above discussion as: solution sets of certain Diophantine equations have additional structure, we may seek for more examples. Another typical example is the case of elliptic curves. For example, the integer solutions to $y^2 = x^3 - Dx$ with $D \in \mathbb{N}$ have a structure of abelian groups. In fact, one can show that this subgroup looks like (some finite group) $\times \mathbb{Z}^r$ for some $r \in \mathbb{Z}_{\geq 0}$. The r is called the rank of this equation, and is an important mathematical invariant.

Now, one may imagine that, as the two equations, albeit different, have a similar structure on their solutions, there should be other aspects of the equations that can be dealt analogously; and precisely this observation inspired mathematicians to transport techniques studying Pell's equation to study the equation $y^2 = x^3 - Dx$ (which of course is considerably more difficult).

1.2. **Definition of groups.**

Definition 1.2.1. A **group** is a pair of a nonempty set G and a binary operation $*$ on G such that

(1) (associativity) $(a * b) * c = a * (b * c)$ for $a, b, c \in G$;
(2) (identity) there exists an element $e \in G$, called the **identity** such that

$$\forall\, a \in G,\ \ a * e = a = e * a;$$

(3) (inverse) for each $a \in G$, there exists $a^{-1} \in G$, called an **inverse** of a such that

$$a * a^{-1} = e = a^{-1} * a.$$

The group G is called **abelian**[1] or **commutative** if $a * b = b * a$ for all $a, b \in G$.
We call $|G|$ the **order** of the group G, (which is possibility infinite).

Remark 1.2.2. Due to the associativity condition, we shall freely remove or add parenthesis when multiplying elements in a group.

Example 1.2.3. (1) $(\mathbb{Z}, +)$ with identity 0. (Remark: many algebra concepts are built to imitate the algebraic structure on $\mathbb{Z}$.)
(2) $(\mathbb{Q}\backslash\{0\}, \cdot)$ with identity 1.
(3) $\mathrm{GL}_n(\mathbb{R}) = \{n \times n$ invertible matrices in $\mathbb{R}\}$.
(4) There can be cases that are not immediately clear to form a group: $(\mathbb{Q}\backslash\{-1\}, *)$ with operation given by $a * b = ab + a + b$. In fact, this is the same as (2) but with numbers shifted by 1.

Example 1.2.4. Fix a positive integer n. Let $\mathbf{Z}_n = \{\bar{0}, \bar{1}, \dots, \overline{n-1}\}$ denotes the full set of residual classes modulo n, where $\bar{i}$ is short for $i \bmod n$. One can add two residual classes, writing $+_n$ for this addition. Explicitly, for $a, b \in \{0, 1, \dots, n-1\}$ set

$$\bar{a} +_n \bar{b} = \begin{cases} \overline{a+b} & \text{if } a+b \leq n-1 \\ \overline{a+b-n} & \text{if } a+b \geq n. \end{cases}$$

This defines a group $(\mathbf{Z}_n, +_n)$. See also Example 0.2.3.

Definition 1.2.5. Let $(G, *)$ and $(H, \circ)$ be groups. Then we may form a new group structure on $G \times H$ with group operation given by

$$(g, h) \star (g', h') := (g * g', h \circ h').$$

This is called the **direct product** of G and H.

[1] This is named after Norwegian mathematician Niels Henrik Abel by Camille Jordan, because Abel had found that the commutativity of the group of a polynomial implies that the roots of the polynomial can be calculated by using radicals.

1.3. **Basic properties of groups.** We list a few basic properties of groups as follows. Let G be a group.

(1) The identity element of a group G is unique.
(If both e and e' are identity elements, $e = e * e' = e'$.)

(2) The inverse of an element $a \in G$ is unique. Better: if an element $b \in G$ satisfies either $b * a = e$ or $a * b = e$, then we have $b = a^{-1}$.
(If $b * a = e$, then we have $b = b * e = b * (a * a^{-1}) = a^{-1}$. The other case is similar.)

(3) $(a^{-1})^{-1} = a$.
(This follows from $a^{-1} \cdot (a^{-1})^{-1} = e$ and the uniqueness of inverse of a^{-1} by (2).)

(4) $(a * b)^{-1} = b^{-1} * a^{-1}$.
(This follows from $(b^{-1} * a^{-1}) * (a * b) = e$ and (2).)

(5) $a * u = a * v$ implies $u = v$. Similarly, $u * b = v * b$ implies $u = v$.
(For the first implication, multiply on the left by a^{-1} gives $a^{-1} * a * u = a^{-1} * a * v$, which further implies $e * u = e * v$, i.e. $u = v$.)

Convention 1.3.1. When writing operations in a group, there are often two conventions:

- **Multiplicative convention**: we typically choose this convention when we do not know whether G is abelian. We write $a \cdot b$ (or simply ab) for the group operation and 1 for the identity. For example:

$$(a_1 a_2 \cdots a_n)^{-1} = a_n^{-1} a_{n-1}^{-1} \cdots a_1^{-1}, \qquad g^n = \underbrace{g \cdot g \cdots g}_{n \text{ times}}.$$

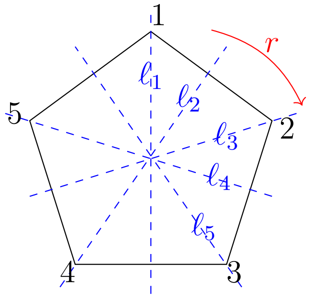
- **Additive convention**: we typically adopt this convention when we want to emphasize that G is abelian. We write $+$ for the group operation, 0 for the identity, and $-a$ for the inverse of a. For example,

$$a + b = b + a, \qquad n \cdot a := \underbrace{a + a + \cdots + a}_{n \text{ times}}.$$

1.4. **Important examples of groups I: Dihedral groups.** A first important example of groups is the dihedral groups

$$D_{2n} = \text{symmetry group of a regular } n\text{-gon}.$$

Since it is non-abelian, we use multiplicative convention for this.



Symmetry of a pentagon

We may list elements of D_{2n} as follows:

$$\begin{aligned} D_{2n} &= \left\{ \begin{array}{l} e = \text{identity},\ r = \text{rotation clockwise } \frac{2\pi}{n},\ r^2,\ \dots,\ r^{n-1} \\ s = s_1 = \text{reflection about } \ell_1,\ s_2 = \text{reflection about } \ell_2,\ \dots,\ s_n \end{array} \right\} \\ &= \left\{ \begin{array}{l} e,\ r,\ r^2,\ \dots,\ r^{n-1} \\ s,\ rs,\ r^2s,\ \dots,\ \dots,\ r^{n-1}s \end{array} \right\}. \end{aligned}$$

Here $rs = s_2$, $r^2s = s_3$, $\dots$. To see this (using the example $n = 5$), we note that vertex $\boxed{1 \xmapsto{s} 1 \xmapsto{r} 2}$. So it must be the reflection about ℓ_2. In particular, $|D_{2n}| = 2n$.

We may rewrite this group in a more efficient form:

$$D_{2n} = \left\langle r, s \;\middle|\; \begin{array}{l} r^n = 1,\ s^2 = 1 \\ srs = r^{-1} \end{array} \right\rangle.$$

This means: D_{2n} consists of the set of words in r, s, r^{-1}, s^{-1} but subject to the given relations. (The relation $srs = r^{-1}$ maybe seen as: drawing this regular n-gon on a piece of paper, then srs is to first flip the paper, then rotation, and then flip back; this is the same as rotating backwards r^{-1}.)

A fun exercise to see is that: $srs = r^{-1}$ implies:

$$sr^is = \underbrace{srs \cdot srs \cdots srs}_{i \text{ copies of } srs} = r^{-1} \cdots r^{-1} = r^{-i}.$$

The first equality comes from inserting many pairs of $s \cdot s^{-1} = s \cdot s$ in the middle of the expression. (This is a standard trick that we will frequently use in group theory.)

Definition 1.4.1. A subset $S = \{s_1, \dots, s_n\}$ of a group G is called a **set of generators** if every element of G can be written as a product of $s_1, \dots, s_n, s_1^{-1}, \dots, s_n^{-1}$.

An equality consisting of generators and their inverses is called a **relation** (e.g. $srs = r^{-1}$)

We write $G = \langle s_1, \dots, s_n \,|\, R_1, R_2, \dots \rangle$ if all relations in G can be deduced from the relations $R_1, R_2, \dots$.[2]

Example 1.4.2. The group $\mathbf{Z}_6 = \langle t \,|\, t^6 = 1 \rangle$.

We may also write $\mathbf{Z}_6 = \langle r, s \,|\, r^3 = s^2 = 1, rs = sr \rangle$ (if r represents $\bar{2}$ and s represents $\bar{3}$).

So there might be many ways to represent the same group using generators and relations.

1.5. **Important examples of groups II: permutation groups.** The second example of groups is the permutation groups or symmetric groups.

Definition 1.5.1. Let Ω be a set. The set

$$S_\Omega := \{\text{bijections } \sigma : \Omega \xrightarrow{\sim} \Omega\}$$

admits a group structure:

- the group operation is composition: $\sigma\tau : \Omega \xrightarrow{\tau} \Omega \xrightarrow{\sigma} \Omega$;
- the identity element is $\mathrm{id} : \Omega \to \Omega$;
- the inverse of the element σ is the inverse map.

[2] This definition is not as rigorous as others, but it is more convenient in writing; we will avoid using it too often.

This S_Ω is called the **symmetry group** or the **permutation group** of Ω.

When $\Omega = \{1, 2, \ldots, n\}$, we write S_n for S_Ω instead.

Notation 1.5.2. There are two ways to represent elements in S_n:

Expression 1: For example, we write

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 7 & 5 & 1 & 3 & 2 & 6 & 4 \end{pmatrix}$$

to mean the bijection that sends $1 \mapsto 7$, $2 \mapsto 5$, $3 \mapsto 1$, $\ldots$ (writing vertically).

We may alternatively express this using a diagram

$$\begin{tikzcd}
& 1 \arrow[dr, mapsto, "\sigma"] & & 2 \arrow[dd, mapsto, bend left, "\sigma"] & \\
3 \arrow[ur, mapsto, "\sigma"] & & 7 \arrow[dl, mapsto, "\sigma"] & & 6 \arrow[loop right, mapsto, "\sigma"] \\
& 4 \arrow[ul, mapsto, "\sigma"] & & 5 \arrow[uu, mapsto, bend left, "\sigma"] &
\end{tikzcd}$$

One sees that σ essentially permutes $1, 7, 4, 3$ in order, and swaps 2 with 5. So we have the following.

Expression 2: We rewrite σ as $(1743)(25)$.

Here for distinct numbers $a_1, \ldots, a_r \in \{1, \ldots, n\}$, we call $(a_1 a_2 \cdots a_r)$ a **cycle**. It represents the permutation of $\{1, \ldots, n\}$ that sends $a_i \mapsto a_{i+1}$ and $a_r \mapsto a_1$ yet keeping all other numbers invariant.

Thus, $\sigma = (1743)(25)$ can be viewed as a composition of two cycles:

$$\begin{tikzcd}
(1743): & 1 \arrow[r, mapsto] & 7 \arrow[r, mapsto] & 4 \arrow[r, mapsto] & 3 \arrow[lll, mapsto, bend right] & \text{and} & (25): & 2 \arrow[r, mapsto, bend left] & 5 \arrow[l, mapsto, bend left]
\end{tikzcd}$$

Writing an element $\sigma \in S_n$ as the product of disjoint cycles is called the **cycle decomposition** of σ.

Properties 1.5.3. (1) S_n is a non-commutative group.

(2) Disjoint cycles commute with each other. This allows us to make effective computation using cycle decompositions. Taking the σ above as an example:

$$\sigma^2 = (1743)^2(25)^2 = (14)(37).$$

$$\sigma^{-1} = (1347)(25).$$

Exercise 1.5.4. Cycles of two elements, namely (ij) are called **transpositions**. Prove the following statements in turn.

(1) The group S_n is generated by all transpositions (ij).
(2) The group S_n is generated by all "adjacent" transpositions $(i, i+1)$.
(3) The group S_n is generated by $\{(12), (123 \cdots n)\}$.

(The relations for S_n with two generators (12) and $(123 \cdots n)$ are somewhat difficult to write down.)

1.6. **Isomorphism of groups.** When there are two groups G and H, we often write e_G and e_H for their identity elements, respectively.

Definition 1.6.1. Two groups $(G, *)$ and $(H, \star)$ are **isomorphic** if there exists a bijection $\phi : G \xrightarrow{\sim} H$ such that, for any $g, h \in G$,

(1) $\phi(g * h) = \phi(g) \star \phi(h)$;
(2) $\phi(e_G) = e_H$;
(3) $\phi(g^{-1}) = \phi(g)^{-1}$.

We write $G \simeq H$ or $\phi : G \xrightarrow{\simeq} H$; such a map ϕ is called an **isomorphism**. (In fact, we will see in the next lecture that condition (1) implies (2) and (3).)

Example 1.6.2. (1) $\exp : (\mathbb{R}, +) \to (\mathbb{R}_{>0}, \cdot)$ is an isomorphism.
(2) The following is an isomorphism.

$$\begin{tikzcd}
\mathbf{Z}_n \arrow[r, "\cong"] & \mu_n = \{\text{all } n\text{th roots of unity in } \mathbb{C}\} \\
a \arrow[r, mapsto] & \zeta_n^a = e^{2\pi i \frac{a}{n}}.
\end{tikzcd}$$

Remark 1.6.3. In group theory, isomorphic groups are considered "same".

$$\boxed{\textbf{Basic question in group theory: classify groups with certain properties.}}$$

For example, all groups of order 6 are either isomorphic to $\mathbf{Z}_6$ or to S_3.

In particular, this says that $D_6 \simeq S_3$ (by identifying the symmetry of a regular triangle with the symmetry of the three vertices). Yet $\mathbf{Z}_6 \not\simeq S_3$ because S_3 is not commutative.

1.7. **Important examples of groups III: Cyclic groups.**

Definition 1.7.1. A group H is called **cyclic** if it can be generated by one element, i.e. there exists $x \in H$, such that $H = \{x^n \mid n \in \mathbb{Z}\}$. Sometimes we write $H = \langle x \rangle$.

The following is clear.

Lemma 1.7.2. *There are two kinds of cyclic groups $H = \langle x \rangle$ (up to isomorphism):*

(1) *If there exists a positive integer n such that $x^n = e$, then take n to be the minimal such number. Then $H = \{1, x, x^2, \dots, x^{n-1}\}$ and H is isomorphic to $\mathbf{Z}_n$ through $\phi : H \xrightarrow{\simeq} \mathbf{Z}_n$ sending $\phi(x^a) = a$. In particular, $|H| = n$.*
(2) *Suppose there does not exist a positive integer n as in (1). Then $H \simeq \mathbb{Z}$ and in particular, $|H| = \infty$.*

Example 1.7.3. The generators of the cyclic group $\mathbf{Z}_n$ are precisely the elements in $\mathbf{Z}_n^\times = \{\bar{a} \mid \gcd(a, n) = 1\}$.

1.8. **Subgroups.** We hope to express the development of the theory of groups parallel to that of vector spaces:

Vector spaces	Groups
Direct sums	Direct products $\surd$
Linear isomorphisms	Isomorphisms $\surd$
Subspaces	Subgroups

Definition 1.8.1. A subset H of a group G is called a **subgroup**, denoted by $H < G$, if

(1) $e \in H$;
(2) for any $a, b \in H$, $ab \in H$;
(3) for any $a \in H$, $a^{-1} \in H$.

There is an alternative definition: a nonempty subset $H \subseteq G$ is a subgroup if and only if

$$\text{for any } a, b \in H \ \Rightarrow \ ab^{-1} \in H.$$

(Note that taking $a = b$ implies $e \in H$; then taking $a = 1$ implies that $b^{-1} \in H$, and finally $a(b^{-1})^{-1} = ab \in H$.)

The subset $\{e\}$ and the entire group G are subgroups of G; they are called the **trivial subgroups** of G.

1.9. **Representing subgroups.**

Definition 1.9.1. Let G be a group and A a subset. Write $\langle A \rangle$ for the **subgroup of** G **generated by** A. Explicitly

$$\big\langle A \big\rangle \;=\; \big\{ a_1^{\epsilon_1} a_2^{\epsilon_2} \cdots a_r^{\epsilon_r} \,\big|\, a_1, \dots, a_r \in A, \epsilon_1, \dots, \epsilon_r \in \{\pm 1\} \big\}$$

It is also the same as the intersection of those subgroups H of G containing A.

Remark 1.9.2. When G abelian and $A = \{a_1, \dots, a_r\}$, we have

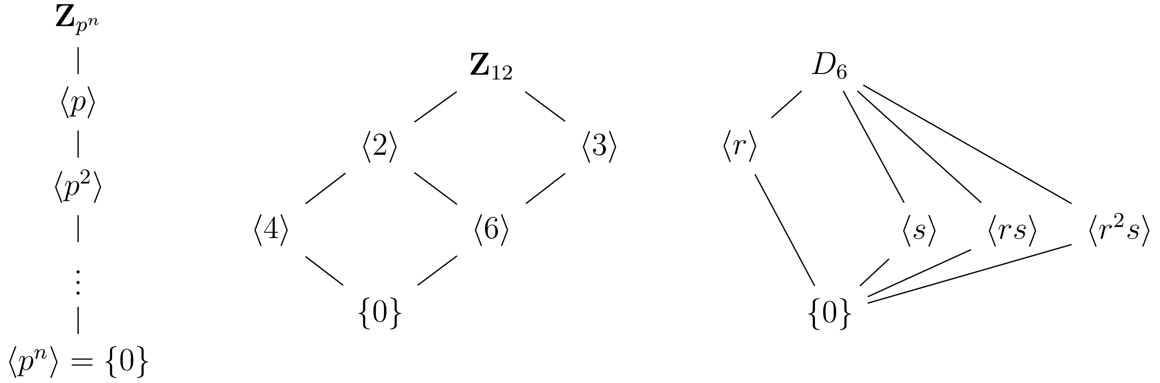
$$\big\langle A \big\rangle = \big\{ a_1^{d_1} \cdots a_r^{d_r} \,\big|\, d_1, \dots, d_r \in \mathbb{Z} \big\}.$$

Definition 1.9.3. Let G be a group and $x \in G$ be an element. Define the **order** of x in G, denoted by $|x|$, as follows

- if there exists integers $a \neq b$ such that $x^a = x^b$, pick a pair $a > b$ with $n = a - b$ minimal, then $x^n = 1$ and $\langle x \rangle = \{1, x, x^2, \dots, x^{n-1}\}$; define $|x| := |\langle x \rangle| = n$;
- if there is no such integers a and b, then $\langle x \rangle = \{1, x, x^2, \dots, x^{-1}, x^{-2}, \dots\} \simeq \mathbb{Z}$; define $|x| = \infty$.

In all cases, $|x| = |\langle x \rangle|$.

1.9.4. *Lattices of subgroups.* Sometimes, it is helpful to enlist subgroups of a group in a diagram encoding their inclusion relations by linking a line between them (with subgroups at below). For example: (p is a prime number)



Extended readings after Section 1

1.1 Matrix groups. Let F be a field (or just think of $F = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ or a finite field $\mathbb{F}_p$ with p a prime number). We may define the **general linear group** with coefficients in F:

$$\mathrm{GL}_n(F) := \big\{A \in \mathrm{M}_{n\times n}(F) \,\big|\, A \text{ is an invertible matrix}\big\}.$$

The group structure is given by matrix multiplication.

This group admits natural (interesting) subgroups.

$$\begin{aligned} B_n(F) &:= \big\{A \in \mathrm{GL}_n(F) \,\big|\, A \text{ is upper triangular}\big\}; \\ N_n(F) &:= \big\{A \in \mathrm{GL}_n(F) \,\big|\, A \text{ is strictly upper triangular}\big\}. \end{aligned}$$

(Strictly upper triangular matrices are the those upper triangular matrices that have all 1 on the diagonal entries.

It is an interesting exercise to see that when $F = \mathbb{F}_p$,

$$|\,\mathrm{GL}_n(\mathbb{F}_p)| = (p^n-1)(p^n-p)(p^n-p^2)\cdots(p^n-p^{n-1}),^3 \quad |B_n(\mathbb{F}_p)| = (p-1)^n p^{(n^2-n)/2}.$$

(We will see later in Lagrange theorem that the order of a subgroup always divides the order of the big group. In this case, $|B_n(\mathbb{F}_p)|$ divides $|\,\mathrm{GL}_n(\mathbb{F}_p)|$, in a quite non-trivial way.)

1.2 The quaternion group. The **quaternion group** Q_8 is the group given by

$$Q_8 = \big\{1, -1, i, -i, j, -j, k, -k\big\}.$$

subject to relations:

$$i\cdot i = j \cdot j = k\cdot k = -1,\ i\cdot j = k,\ j\cdot k = i,\ k \cdot i = j.$$

The name comes from the *quaternion algebra* (or the *Hamiltonian algebra*):

$$\mathbb{H} := \mathbb{R} \oplus \mathbb{R}\cdot i \oplus \mathbb{R}\cdot j \oplus \mathbb{R}\cdot k.$$

subject to the above same relation. This $\mathbb{H}$ is a *non-commutative* ring, in which every nonzero element admits an inverse. (Such rings are called **division rings** or **skew fields**.)

The quaternion group embeds into $\mathbb{H}^\times$ as a subgroup.

1.3 Subgroups of cyclic groups. We show that every subgroup of a cyclic group is cyclic. More precisely, let $G = \langle x\rangle$ be a cyclic group. Assume that n is the minimal positive integer such that $x^n = e$ if such n exists, and $n = 0$ otherwise.

Let H be a subgroup of G. Assume that $H \neq \{e_G\}$ (otherwise the statement is trivial.) Let m be the minimal positive integer such that $x^m \in H$. We claim that $H = \langle x^m\rangle$, namely, every element of H is of the form x^{mr} for some $r \in \mathbb{Z}$.

Suppose not, say $x^a \in H$ with $a \in \mathbb{Z}$ not divisible by m. Write $a = mr + s$ with $r, s \in \mathbb{Z}$ and $s \in \{0, \dots, m-1\}$. Then $x^s = x^a(x^m)^{-r} \in H$, contradicting with the minimality of m.

This completes the proof of: every subgroup of a cyclic group is cyclic. More precisely,

(1) When $H \simeq \mathbf{Z}_n$, any subgroup is generated by $\langle x^m\rangle$ for some minimal m. It is not hard to see that $x^{\gcd\{m,n\}} \in H$; so such m must be a divisor of n. Any for any divisor m of n, $\langle x^m\rangle$ is a subgroup of $\mathbf{Z}_n$, of order $\frac{n}{m}$.

(2) When $H \simeq \mathbb{Z}$, any subgroup is of the form $\langle x^m\rangle$ for some $m \in \mathbb{N}$; it is cyclic and infinite.

[3] One can prove this as follows: the first column of the matrix must be nonzero, so there are $p^n - 1$ choices; then the second column cannot be a scalar multiple of the first column, so there are $p^n - p$ choices; …

2. Cosets, Lagrange theorem, quotient groups

We continue with parallel development of vector spaces versus groups.

Vector spaces	Groups
Direct sums	Direct products $\surd$
Linear isomorphisms	Isomorphisms $\surd$
Subspaces	Subgroups $\surd$
Affine subsets $v + W$	Cosets
Quotient spaces	Quotient groups
Linear maps	Homomorphisms

2.1. Cosets.

Cosets may be viewed as analogues of affine subsets in linear algebra.

Definition 2.1.1. Let H be a subgroup of G. A **left coset** is a set of the form (for some $g \in G$)

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

Note that g always belongs to gH, so we say that g is a **representative** for the coset gH.

In particular, if $g \in H$, then $gH = H$.

A **right coset** is a set of the form (for some $g \in G$)

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

Remark 2.1.2. Occasionally, people may abbreviate left cosets to simply cosets. (We will mostly work with left cosets throughout, but similar statement should also hold for right cosets.)

When G is abelian, left cosets and right cosets are the same.

Convention 2.1.3. In what follows, we will frequently use the notation of gH or HK for an element $g \in G$ and subsets $H, K \subseteq G$. By writing this, we meant just simply element-wise multiplication

$$gH = \{gh \mid h \in H\} \quad \text{and} \quad HK = \{hk \mid h \in H, k \in K\}.$$

This somewhat simplifies our discussion, as we frequently encounter the situation that $g_1H = g_2H$ with $g_1 \neq g_2$. But viewing this as a subset circumvent the discussion of choosing a coset representative.

Proposition 2.1.4. *Two (left) cosets g_1H and g_2H are*

- *either* equal *(which is equivalent to $g_1^{-1}g_2 \in H$);*
- *or* disjoint *(which is equivalent to $g_1^{-1}g_2 \notin H$).*

In particular, G is the disjoint *union of left cosets for H.*

Proof. We will prove that

$$\begin{tikzcd} g_1H \cap g_2H \neq \emptyset \arrow[r, Rightarrow, "(B)"] & g_1^{-1}g_2 \in H \arrow[r, Rightarrow, "(A)"] & g_1H = g_2H \arrow[ll, Rightarrow, bend right, "\text{obvious}"'] \end{tikzcd}$$

(A) If $g_1^{-1}g_2 \in H$, then $g_1H = g_1(g_1^{-1}g_2 \cdot H) = g_2H$ (as for any element $h \in H$, $h \cdot H = H$ as sets).

(B) If $g_1H \cap g_2H \neq \emptyset$, say $x = g_1h_1 = g_2h_2$ for $h_1, h_2 \in H$. Then $g_1 = xh_1^{-1}$ and $g_2 = xh_2^{-1}$. So

$$g_1^{-1}g_2 = (xh_1^{-1})^{-1} \cdot xh_2^{-1} = h_1x^{-1}xh_2^{-1} = h_1h_2^{-1} \in H.$$

Finally, the equivalence above shows the equivalence of first bullet point of the proposition, and also at the same time the equivalence to the negative of the second bullet point. □

2.2. **Lagrange Theorem.** The first big theorem in group theory is Lagrange's theorem.

Definition 2.2.1. Write $G/H := \{gH \mid g \in G\}$ for the set of left cosets. Similarly, $H\backslash G := \{Hg \mid g \in G\}$ for the right cosets.

The above proposition says $G = \coprod_{gH \in G/H} gH$ is a disjoint union.

We call $[G : H] := |G/H|$ the **index** of H as a subgroup of G (possibly infinite).

Theorem 2.2.2 (Lagrange)**.** *If G is a finite group and $H < G$ is a subgroup, then $|H|$ divides $|G|$.*

Proof. As each coset for H has exactly $|H|$ elements, we have $|G| = [G : H] \cdot |H|$. □

Corollary 2.2.3. *(1) If G is a finite group, then $|x|$ divides $|G|$.*

(2) For every element $x \in G$, $x^{|G|} = e$.

Proof. (1) This follows from Lagrange theorem because $|x| = |\langle x\rangle|$ and $\langle x\rangle$ is a subgroup of G.

(2) follows from (1) immediately. □

Example 2.2.4. Following Example 0.2.3, fix a positive integer n. Consider $G = \mathbb{Z}_n^\times := \{\bar{a} = a \bmod n \mid \gcd(a, n) = 1\}$, the group of modulo n residual classes that are coprime to n. The operation for the group G is the multiplication. We know that $|G| = \varphi(n)$ is the Euler function.

Then Lagrange theorem for G says that, if $\gcd(a, n) = 1$, $\bar{a}^{\varphi(n)} = \bar{1}$; or equivalently,

$$\gcd(a, n) = 1 \quad \Longrightarrow \quad a^{\varphi(n)} \equiv 1 \pmod{n}.$$

This is known as the Euler's theorem generalizing Fermat's Little Theorem.

Corollary 2.2.5. *If a group G has p elements with p a prime, then G is cyclic (and in particular abelian).*

Proof. Take an element $x \in G$ such that $x \neq e$. Then $|x|$ divides $|G| = p$. Yet $|x| \neq 1$; so $|x| = p$, i.e. $|\langle x\rangle| = p$. So $G = \langle x\rangle$ is cyclic. □

2.3. **Conjugation, normal subgroups, and quotient groups.**

Definition 2.3.1. Let G be a group and $a, g \in G$. We call gag^{-1} the **conjugate of** a **by** g.

Lemma 2.3.2. *If H is a subgroup of G and $g \in G$, then $gHg^{-1} := \{ghg^{-1} \mid h \in H\}$ is a subgroup, called the* **conjugate of** H **by** g.

Proof. Given $gag^{-1}, gbg^{-1} \in gHg^{-1}$,

$$(gag^{-1})(gbg^{-1})^{-1} = gag^{-1} \cdot gb^{-1}g^{-1} = gab^{-1}g^{-1} \in gHg^{-1}.$$

So gHg^{-1} is a subgroup of G. □

Definition 2.3.3. A subgroup $H \leqslant G$ is **normal** if for any $g \in G$, $H = gHg^{-1}$; namely, all conjugates of H are just H itself. Note that this condition is also equivalent to $gH = Hg$ (as subsets) for any $g \in G$, namely the left coset of g is the same as right coset of g for each g.

We write $H \trianglelefteq G$ to denote normal subgroups. For example, $\{1\} \trianglelefteq G$ and $G \trianglelefteq G$.

Definition 2.3.4. Let $H \trianglelefteq G$ be a normal subgroup. For $a, b \in G$, we have

$$aH \cdot bH := \{k\ell \,|\, k \in aH,\ \ell \in bH\} = abH \cdot H = abH$$

as subsets of G. (Note that, we used that $Hb = bH$ in the second equality; this avoids the discussion of whether the product is well-defined.) This defines a group structure on G/H. The identity is $eH = H$; and the inverse of aH is $a^{-1}H$.

We call G/H the **quotient group** or the **factor group** of G by H.

Example 2.3.5. (1) Every subgroup of an abelian group is normal, because $gHg^{-1} = H$ automatically holds.

(2) A positive integer n defines a (normal) subgroup of $(\mathbb{Z}, +)$ given by $\langle n \rangle = \{x \in \mathbb{Z} \,|\, n \text{ divides } x\}$. The quotient group $\mathbb{Z}/\langle n \rangle$ has elements $a + \langle n \rangle$ for $a = 0, 1, \ldots, n-1$. Then we have a natural isomorphism

$$\begin{tikzcd} \mathbb{Z}/\langle n \rangle \arrow[r, "\cong"] & \mathbf{Z}_n \\ a + \langle n \rangle \arrow[r, mapsto] & \bar{a} \end{tikzcd}$$

We often write this quotient as $\mathbb{Z}/n\mathbb{Z}$ and thus $\mathbf{Z}_n \cong \mathbb{Z}/n\mathbb{Z}$ (this may be viewed as the definition of $\mathbf{Z}_n$).

2.4. **Some technical results.**

Proposition 2.4.1. *Let H and K be subgroups of a group G. Define $HK := \{hk \,|\, h \in H,\ k \in K\}$. When G is finite, we have*

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

Proof. By Proposition 2.1.4, HK is a disjoint union of left cosets of K, namely

$$HK = h_1K \sqcup h_2K \sqcup \cdots \sqcup h_mK.$$

We claim that for the same $h_1, \ldots, h_m$,

$$H = h_1(H \cap K) \sqcup \cdots \sqcup h_m(H \cap K).$$

The claim implies the proposition as

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

Now, we prove the claim: for every $h, h' \in H$,

$$hK = h'K \;\Leftrightarrow\; h^{-1}h' \in K \;\Leftrightarrow\; h^{-1}h' \in H \cap K \;\Leftrightarrow\; h(H \cap K) = h'(H \cap K). \tag{2.4.1.1}$$

So we deduce that

$$\begin{aligned} HK &= \bigcup_{h\in H} hK = h_1K \sqcup \cdots \sqcup h_mK \\ H &= \bigcup_{h\in H} h(H\cap K) = h_1(H\cap K) \sqcup \cdots \sqcup h_m(H\cap K). \end{aligned}$$

Here, the equalities mean to first write tautologically HK as the union of all left K-cosets (parameterized by $h \in H$) and in a parallel way write H as the union of all left $H \cap K$-cosets (parameterized by $h \in H$). Then (2.4.1.1) tells us that two left K-cosets are the same if and only if the corresponding left $H \cap K$-cosets are the same. Thus, in the next step writing union of cosets into disjoint union of cosets, both equalities may use the same set of representatives. □

A more "fancy" proof. In fact, we prove a much stronger statement: there is a canonical bijection between coset spaces:

$$\begin{aligned} \varphi : H/(H\cap K) &\longrightarrow HK/K \\ h(H\cap K) &\longmapsto hK. \end{aligned}$$

At first glance, one needs to verify that φ is well-defined; but we may rewrite φ as $\varphi\big(h(H\cap K)\big) := h(H\cap K)K$ as product of subsets of G in the sense of Convention 2.1.3. But then $h(H\cap K)K = hK$ for trivial reasons. So φ is well-defined.

Now we note that φ is clearly surjective, as every cosets in HK/K takes the form of hkK with $h \in H$ and $k \in K$, which is the same as $hK = \varphi(h(H\cap K))$.

Finally, we show that φ is injective, i.e. for two cosets $h_1(H\cap K)$ and $h_2(H\cap K)$ with $h_1, h_2 \in H$, if they have the same image under φ, then $h_1K = h_2K$; then $h_2^{-1}h_1 \in K$. But $h_2^{-1}h_1 \in H$ for trivial reasons, so $h_2^{-1}h_1 \in H\cap K$ and thus $h_1(H\cap K) = h_2(H\cap K)$.

This completes the proof of that φ is bijective. The proposition is clear from this. □

Caveat 2.4.2. The set HK above need not be a group. For example, in $G = D_6$, $H = \langle s_1\rangle$ and $K = \langle s_2\rangle$, then $|HK| = |H|\cdot|K| = 4$. But HK cannot be a subgroup of D_6 because $4 \nmid 6$.

Lemma 2.4.3. *Let H and K be subgroups of G. If $HK = KH$ as sets (meaning every product kh with $k \in K$ and $h \in H$ can be rewritten as $h'k'$ with $k' \in K'$ and $h' \in H'$), then HK is a subgroup of G.*

In particular, if K is a normal subgroup, then $hK = Kh$ for any $h \in H$, and thus $HK = KH$ is a subgroup of G.

Proof. For $h_1, h_2 \in H$ and $k_1, k_2 \in K$, we need to check that

$$h_1k_1 \cdot (h_2k_2)^{-1} = h_1k_1k_2^{-1}h_2^{-1}$$

belongs to HK. Yet the condition says that $k_1k_2^{-1}\cdot h_2^{-1} = h'k'$ for some $h' \in H'$ and $k' \in K'$. Thus

$$h_1k_1 \cdot (h_2k_2)^{-1} = h_1k_1k_2^{-1}h_2^{-1} = h_1h'k' \in HK.$$

□

Lemma 2.4.4. *If H and K are both normal subgroups of G, then HK is also a normal subgroup of G.*

Proof. We have checked that HK is a subgroup. For any $g \in G$, we check

$$gHK = HgK = HKg.$$

So HK is a normal subgroup of G. □

2.5. **Group homomorphisms.** The concept of group homomorphisms may be viewed as a way to relate two groups.

Definition 2.5.1. Let $(G, *)$ and $(H, \star)$ be two groups. A map $\phi : G \to H$ is called a **homomorphism** if for any $x, y \in G$, we have

$$\phi(x * y) = \phi(x) \star \phi(y).$$

Remark 2.5.2. (1) For a group homomorphism $\phi : G \to H$,

$$\phi(e_G) = e_H, \quad \text{and} \quad \phi(g^{-1}) = \phi(g)^{-1}.$$

Indeed, $\phi(e_G) = \phi(e_G * e_G) = \phi(e_G) \star \phi(e_G)$. Thus $e_H = \phi(e_G)$ by cancellation. For each $g \in G$, we have $e_H = \phi(e_G) = \phi(g * g^{-1}) = \phi(g) \star \phi(g^{-1})$. So $\phi(g^{-1}) = \phi(g)^{-1}$.

(2) A bijective group homomorphism is a group isomorphism.

Example 2.5.3. (1) $\phi : \mathbb{Z} \to \mathbf{Z}_n$ given by $\phi(m) = m \bmod n$ is a homomorphism.

(2) When H is a normal subgroup of G, there is a natural *surjective* homomorphism

$$\begin{tikzcd} G \arrow[r, "\pi"] & G/H \\ a \arrow[r, mapsto] & \pi(a) = aH. \end{tikzcd}$$

For this, we check that $\pi(ab) = abH = aH \cdot bH = \pi(a) \cdot \pi(b)$.

Lemma 2.5.4. *If $\phi : (G, *) \to (H, \star)$ and $\psi : (H, \star) \to (K, \bullet)$ are two homomorphisms, then the composition $\psi \circ \phi : (G, *) \to (K, \bullet)$ is a homomorphism.*

Proof. We simply check that for $x, y \in G$, we have

$$\psi \circ \phi(x * y) = \psi(\phi(x) \star \phi(y)) = \psi(\phi(x)) \bullet \psi(\phi(y)).$$

□

2.6. **Kernel of a group homomorphism.**

Definition 2.6.1. For a homomorphism $\phi : G \to H$ of groups, the **kernel** is

$$\ker \phi := \{g \in G \mid \phi(g) = e_H\}.$$

Lemma 2.6.2. *Let $\phi : G \to H$ be a group homomorphism.*

(1) *The image $\phi(G)$ is a subgroup of H.*

(2) *The kernel $\ker \phi$ is a normal subgroup of G.*

Proof. (1) It follows from that $\phi(g_1)\phi(g_2)^{-1} = \phi(g_1)\phi(g_2^{-1}) = \phi(g_1 g_2^{-1}) \in \phi(G)$.

(2) If $g_1, g_2 \in \ker \phi$, then

$$\phi(g_1 g_2^{-1}) = \phi(g_1)\phi(g_2)^{-1} = e_H e_H^{-1} = e_H.$$

So $g_1 g_2^{-1} \in \ker \phi$. Thus $\ker \phi$ is a subgroup.

For any $g' \in G$ and any $g \in \ker \phi$,

$$\phi(g' g g'^{-1}) = \phi(g')\phi(g)\phi(g')^{-1} = \phi(g') e_H \phi(g')^{-1} = e_H.$$

So $g' g g'^{-1} \in \ker \phi$. This implies that $\ker \phi$ is a normal subgroup of G. □

The triviality of the kernel indicates the injectivity of a homomorphism.

Lemma 2.6.3. *A homomorphism $\phi : G \to H$ of groups is injective if and only if* $\ker \phi = \{e_G\}$.

Proof. The injectivity $\Rightarrow \ker \phi = \{e_G\}$ is clear as $\phi(e_G) = e_H$.

Conversely, suppose that $\ker \phi = \{e_G\}$, we need to show that ϕ is injective.

Suppose that $\phi(g_1) = \phi(g_2)$ for some $g_1, g_2 \in G$. Then

$$\phi(g_1 g_2^{-1}) = \phi(g_1)\phi(g_2)^{-1} = e_H.$$

Thus $g_1 g_2^{-1} \in \ker \phi = \{e_G\}$. So $g_1 g_2^{-1} = e_G$, and thus $g_1 = g_2$. This proves that ϕ is injective. $\square$

Extended readings after Section 2

2.1 Describing a homomorphism. If a group G is given by generators and relations as $\langle s_1, \dots, s_n \,|\, R_1, R_2, \dots \rangle$, to specify a homomorphism $\phi : G \to H$ to another group H is equivalent to specify the images $\phi(s_1), \dots, \phi(s_n)$ of the generators (as elements in H), such that the relation $\phi(R_j)$ still hold.

Instead of giving a proof of the above statement, we illustrate it through an example.

Let us compute all homomorphisms $\phi : D_{2n} \to \mathbb{C}^\times$. Recall that $D_{2n} = \big\langle r, s \,\big|\, r^n = s^2 = 1, srs = r^{-1} \big\rangle$. To describe such a homomorphism ϕ, we just need to give two numbers $\phi(r), \phi(s) \in \mathbb{C}^\times$ such that

$$\phi(r)^n = \phi(s)^2 = 1 \quad \text{and} \quad \phi(s)\phi(r)\phi(s) = \phi(r)^{-1}.$$

The second relation says $\phi(r)\phi(s)^2 = \phi(r)^{-1}$. Yet $\phi(s)^2 = 1$, so $\phi(r)^2 = 1$. From this, we can easily deduce that

- when n is odd, $\phi(s) \in \{\pm 1\}$ and $\phi(r) = 1$;
- when n is even, $\phi(s) \in \{\pm 1\}$ and $\phi(r) \in \{\pm 1\}$.

Continued with this, we may try to understand index 2 subgroups of D_{2n}. It is an exercise to show that every subgroup of index 2 is normal. So we have a bijection

$$\begin{tikzcd}
\big\{\text{index 2 subgroups}\big\} \arrow[r, leftrightarrow] & \big\{\text{nontrivial homomorphisms } \phi : D_{2n} \to \{\pm 1\}\big\} \\
H \arrow[r, mapsto] & \phi : G \twoheadrightarrow G/H \simeq \{\pm 1\} \\
\ker(\phi) & \pi \arrow[l, mapsto]
\end{tikzcd}$$

This allows us to enlist all index 2 subgroups as follows.

- When n is odd, the only nontrivial homomorphism ϕ is given so that $\phi(s) = -1$ and $\phi(r) = 1$. So $\ker(\phi) = \langle r \rangle$ is the only index 2 subgroup of D_{2n}.
- When n is even, there are three nontrivial homomorphisms

$\phi(s)$	$\phi(r)$	$\ker(\phi)$
1	-1	$\langle s, r^2 \rangle$
-1	1	$\langle r \rangle$
-1	-1	$\langle sr, r^2 \rangle$

3. Isomorphism theorems, composition series, statement of Hölder theorem

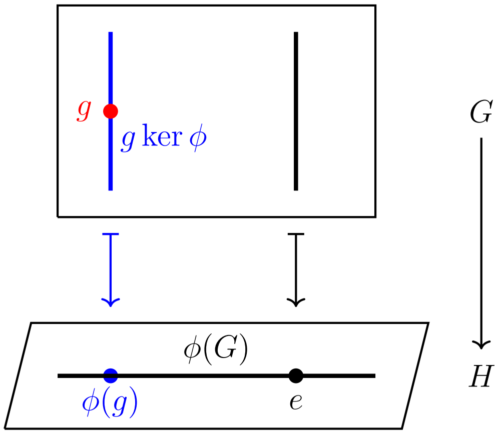
In this lecture, we will introduce a series of isomorphism theorems to relate the source and target groups of a homomorphism. After this, we explain Hölder's program to classify all groups.

3.1. **The isomorphism theorems.**

Theorem 3.1.1 (The first isomorphism theorem)**.** *If $\phi : G \to H$ is a homomorphism of groups, then $\ker\phi \trianglelefteq G$ and*

$$G/\ker\phi \cong \phi(G).$$

Proof. It may help to visualize the situation of the theorem as follows:



We define the needed map

$$\begin{aligned} \psi : G/\ker\phi &\longrightarrow \phi(G) \\ g\ker\phi &\longmapsto \phi(g). \end{aligned}$$

We need to prove the following:

(1) ψ is well-defined. Suppose $g_1\ker\phi = g_2\ker\phi$. Then $g_2^{-1}g_1 \in \ker\phi$, and thus

$$\phi(g_1) = \phi(g_2 \cdot g_2^{-1}g_1) = \phi(g_2)\cdot\phi(g_2^{-1}g_1) = \phi(g_2)\cdot e_H = \phi(g_2).$$

(2) ψ is surjective. This is because every element of $\phi(G)$ takes the form of $f(g)$; and it is the image of $g\ker\phi$.

(3) ψ is injective. It is enough to check that $\ker\psi = \{\ker\phi\}$.

This is because if $\psi(g\ker\phi) = \phi(g) = e_H$, then $g \in \ker\phi$. Thus $g\ker\phi = \ker\phi$. So $\ker\psi = \{\ker\phi\}$.

(4) ψ is a homomorphism. One checks this as:

$$\begin{tikzcd}
\psi\big(g_1\ker\phi \cdot g_2\ker\phi\big) \arrow[r, equal] & \psi(g_1g_2\ker\phi) \arrow[r, equal] & \phi(g_1g_2) \arrow[d, equal] \\
\psi(g_1\ker\phi)\cdot\psi(g_2\ker\phi) \arrow[rr, equal] & & \phi(g_1)\psi(g_2).
\end{tikzcd}$$

□

Theorem 3.1.2 (The second isomorphism theorem)**.** *Let G be a group, and let $A \leq G$ be a subgroup and $B \trianglelefteq G$ a normal subgroup. Then AB is a subgroup of G, $B \trianglelefteq AB$, $A \cap B \trianglelefteq A$, and*

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

Proof. We have shown in Lemma 2.4.3 that $AB \leq G$.

We first prove that $B \trianglelefteq AB$: indeed, since $B \trianglelefteq G$, we have for any $g \in G$, $gBg^{-1} = B$. The same equality certainly holds for $g \in AB \subseteq G$. So $B \trianglelefteq AB$ and hence the quotient AB/B makes sense.

Now define a homomorphism

$$\begin{tikzcd}
\phi : A \arrow[r] & AB \arrow[r, two heads] & AB/B \\
a \arrow[r, maps to] & a \arrow[r, maps to] & aB.
\end{tikzcd}$$

We will show:

- ϕ is clearly surjective, because $abB = aB = \phi(a)$.
- $\ker \phi = \{a \in A \,|\, aB = B\}$. (the condition $aB = B$ implies $a \in B$.) So $\ker \phi = A \cap B$. In particular, as a kernel, $A \cap B$ is a normal subgroup of A.

Now, by the first isomorphism theorem, we deduce

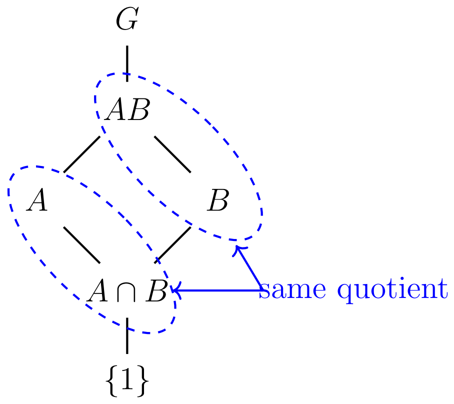
$$A/(A \cap B) \xrightarrow{\cong} AB/B.$$

□

Remark 3.1.3. A naive way to prove two groups G and H are isomorphic is to establish a bijection between G and H and show that this is a homomorphism. A "more advanced way" is to first establish a homomorphism between some groups G' and H' "closely related" to G and H. Then we use isomorphism theorems to try to relate G with H.

In the particular case of proving some statement like $G/H \cong G'$. One may first construct a *surjective* homomorphism $\phi : G \to G'$ and compute the kernel of ϕ to be H.

Remark 3.1.4. We have the following diagram of subgroups.



Remark 3.1.5. The statement of the theorem is slightly weaker than the one from the book. Instead of requiring B to be a normal subgroup of G, it suffices to require that A normalizes B, i.e. $\forall a \in A$, $aBa^{-1} = B$. The only place where we need some modification is where we prove $B \trianglelefteq AB$. Indeed, given $a \in A$ and $b \in B$, we have

$$abB(ab)^{-1} = abBb^{-1}a^{-1} = aBa^{-1} = B.$$

Theorem 3.1.6 (The third isomorphism theorem)**.** *Let G be a group and H and K be normal subgroups with $H \leq K$. Then $K/H \trianglelefteq G/H$, and*

$$(G/H)\big/(K/H) \cong G/K.$$

(If we denote the quotient by H using a bar, then this says that $\overline{G}/\overline{K} \cong G/K$.

One can alternatively write this in terms of a diagram:

$$\begin{tikzcd}
G \arrow[d, dash] \arrow[d, dash, bend left, "G/K\cong(G/H)/(K/H)"] \arrow[dd, dash, bend right=60, "G/H"'] \\
K \arrow[d, dash] \arrow[d, dash, bend left, "K/H"] \\
H
\end{tikzcd}$$

Proof. Consider the map

$$\begin{aligned} \phi : G/H &\longrightarrow G/K \\ gH &\longmapsto gK. \end{aligned}$$

- ϕ is well-defined. We can simply redefine ϕ as $\phi(gH) = gH \cdot K = gK$ as product of subsets of G.
- ϕ is a homomorphism. This is because

$$\begin{tikzcd}
\phi(g_1H \cdot g_2H) \arrow[r, equal] & \phi(g_1g_2H) \arrow[r, equal] & g_1g_2K \arrow[d, equal] \\
\phi(g_1H) \cdot \phi(g_2H) \arrow[rr, equal] & & g_1K \cdot g_2K.
\end{tikzcd}$$

- ϕ is surjective. This is clear.
- $\ker\phi$ is equal to

$$\ker\phi = \big\{gH \,\big|\, gK = K\big\} = \big\{gH \,|\, g \in K\big\} = K/H.$$

  In particular, (as a kernel), K/H is a normal subgroup of G/H.

Using the first isomorphism theorem, we deduce that

$$(G/H)\big/(K/H) \cong G/K.$$

□

Theorem 3.1.7 (The fourth isomorphism theorem / Lattice isomorphism theorem)**.** *Let G be a group and $N \trianglelefteq G$ a normal subgroup. Then there is a bijection*

$$\begin{aligned} \big\{\textit{subgroups of } G \textit{ containing } N\big\} &\longleftrightarrow \big\{\textit{subgroups of } G/N\big\} \\ A &\longmapsto A/N \\ \pi^{-1}(\overline{A}) &\longleftarrow\!\shortmid\ \overline{A} \end{aligned}$$

where $\pi : G \to G/N$ is the natural projection.

This bijection preserves

- *inclusions of groups,*
- *index of subgroups,*

- *intersections,*
- *normality of subgroups, and*
- *quotients of subgroups.*

Visually, we have

$$\textit{Lattice of subgroups of } G \textit{ containning } N \longleftrightarrow \textit{Lattice of subgroups of } G/N.$$

3.2. **Universal property of quotient groups.** Consider the following situation $\phi : G \to H$ is a group homomorphism and let $N \trianglelefteq G$ be a normal subgroup. We hope to define

$$\begin{aligned} \Phi : G/N &\longrightarrow H \\ gN &\longmapsto \phi(g). \end{aligned}$$

Lemma 3.2.1. *Such map Φ is well-defined if and only if $N \subseteq \ker \phi$. In this case, Φ is a homomorphism.*

Proof. This is because if $g_1N = g_2N$, then $g_1 = g_2n$ for some $n \in N$. Thus we need to see whether

$$\phi(g_1) = \phi(g_2n) = \phi(g_2)\phi(n) \overset{?}{=} \phi(g_2).$$

This happens if and only if $\phi(n) = e_H$ (i.e. $n \in \ker \phi$). From this, it is clear that Φ is well-defined if and only if $N \subseteq \ker \phi$ (as n may be taken to be any element in N.)

Moreover, it is clear that in this case Φ is a homomorphism. □

Example 3.2.2. We explain how this lemma is used. Consider a homomorphism $\phi : \mathbb{Z} \to \mathbb{C}^\times$ (such a homomorphism is in fact determined by the value $\lambda := \phi(1) \in \mathbb{C}^\times$. Then we ask the question: which ϕ induces a well-defined homomorphism $\mathbb{Z}/\langle n \rangle \to \mathbb{C}^\times$? For this, we need $\phi(\langle n \rangle) = 1$. This is equivalent to requiring $\phi(n) = 1$, or in other words, $\lambda^n = 1$.

Notation 3.2.3. When $N \subseteq \ker \phi$, we say that $\phi : G \to H$ **factors through** G/N, graphically,

$$\begin{tikzcd}
G \arrow[r, "\phi"] \arrow[d, two heads, "\pi"'] & H & & g \arrow[r, mapsto] \arrow[d, mapsto] & \phi(g) \\
G/N \arrow[ur, dashed, "\Phi"'] & & & gN. \arrow[ur, dashed, mapsto] &
\end{tikzcd}$$

this says that there exists a unique Φ that **makes the diagram commute**.

Remark 3.2.4. This is the first time that we introduce a diagrammatic description of a statement. In future study of abstract algebra or commutative algebra, we will, with some frequency, find that such diagrammatic description helpful to clarify the situation.

Remark 3.2.5. One may view the lemma above as: when we consider the homomorphism $\phi : G \to H$, one may to do this in two steps: first "group together" the information in N, and then map to H.

Another way to think of this is that: if we consider all possible homomorphisms out of G to some group H such that the kernel contains N, this is equivalent to consider homomorphism (first to G/N and then) out of G/N to H. This is the "universal property" of the quotient G/N. (In other words, if we want to consider homomorphisms out of G whose kernel contains N, it suffices to look at G/N.)

3.3. **Hölder's program.**

$$\boxed{\text{Ultimate goal of group theorists: classify all finite groups.}}$$

One observation we get from above is that, if $N \lhd G$ is a proper normal subgroup (i.e. $N \neq \{1\}$, G). Then roughly, we may obtain some information of G from that of N and of G/N.

Definition 3.3.1. A (finite or infinite) group G is called **simple** if $|G| > 1$ and the only normal subgroups of G are $\{1\}$ and G.

Example 3.3.2. (1) $\mathbf{Z}_p$ for a prime number p. (This is all abelian simple groups.)
(2) Alternating group A_n for $n \geq 5$ (a subgroup of S_n we introduce later).
(3) There are infinite simple groups, but not so easy to define.

So the **Hölder program** consists of two steps:

- Step I: classify all finite simple groups;
- Step II: Find all ways of "putting simple groups together" to form other groups.

The following is considered the most important achievement of group theory.

Theorem 3.3.3 (Classification of finite simple groups)**.** *Every finite simple group is isomorphic to one in*

- 18 *(infinite) families of simple groups, or*
- 26 *sporadic simple groups.*

The list of family of finite groups includes

- $\mathbf{Z}_p$ with p a prime;
- A_n $(n \geq 5)$;
- $\mathrm{PSL}_n(\mathbb{F}) = \mathrm{SL}_n(\mathbb{F})/Z(\mathrm{SL}_n(\mathbb{F}))$ with $n \geq 2$ and $\mathbb{F}$ a finite field (e.g. $\mathbb{F}_p$). (Here $Z(\mathrm{SL}_n(\mathbb{F}))$ is the scalar matrices with coefficients in $\mathbb{F}^\times$ and whose determinant is 1.)

There are other lists of finite groups mostly associated to a family of "Lie groups of finite type".

For the interest of readers, we only on mention the following.

Theorem 3.3.4 (Feit–Thompson theorem)**.** *If G is a simple group of odd order, then $G \cong \mathbf{Z}_p$ for some odd prime p.*

Remark 3.3.5. In fact, the original Feit–Thompson theorem states that every finite group of odd order is solvable! (See the definition of solvable groups below in Definition 3.4.4.)

Certainly, in this abstract algebra course, we will only touch the very basics of group theory. We hope to learn some tools that frequently appears in applications to future questions in algebra.

3.4. **Composition series.** Inspired by the Hölder's program, we make the following.

Definition 3.4.1. In a group G, a sequence of subgroups

$$\{1\} = N_0 \leq N_1 \leq \cdots \leq N_k = G$$

is called a **composition series** if $N_{i-1} \trianglelefteq N_i$ and N_i/N_{i-1} is a simple group for each $1 \leq i \leq k$.

In this case, the factor groups N_i/N_{i-1} are called **composition factors** or **Jordan–Hölder factors** of G.

Example 3.4.2. For the dihedral group $D_8 = \langle r, s \,|\, r^4 = s^2 = 1,\ srs = r^{-1}\rangle$, the following are two composition series (and there are more):

- $\{1\} \lhd \langle s\rangle \lhd \langle s, r^2\rangle \lhd D_8$,
- $\{1\} \lhd \langle r^2\rangle \lhd \langle r\rangle \lhd D_8$.

Theorem 3.4.3 (Jordan–Hölder)**.** *Let G be a nontrivial finite group. Then*

1. *G has a composition series, and*
2. *the composition factors are unique up to permutation, i.e. if we have two composition series*
$$\{1\} = A_0 \lhd A_1 \lhd \cdots \lhd A_m = G \quad \text{and} \quad \{1\} = B_0 \lhd B_1 \lhd \cdots \lhd B_n = G,$$
then $m = n$, and there exists a bijection $\sigma : \{1, \ldots, m\} \to \{1, \ldots, n = m\}$ such that, for $i = 1, \ldots, m$,
$$A_i/A_{i-1} \simeq B_{\sigma(i)}/B_{\sigma(i)-1}.$$

Proof of (1). This is because if G is simple, then $\{1\} \lhd G$ itself forms a composition series. If G has a nontrivial normal subgroup N, then we may immediately reduce to N and G/N as follows: writing $\pi : G \to G/N$ and giving composition series
$$\{1\} = C_0 \lhd C_1 \lhd \cdots \lhd C_r = N \quad \text{and} \quad \{N\} = D_0 \lhd D_1 \lhd \cdots \lhd D_s = G/N,$$
we may "combine" them using the Fourth Isomorphism Theorem as
$$\{1\} = C_0 \lhd C_1 \lhd \cdots \lhd C_r = N = \pi^{-1}(D_0) \lhd \pi^{-1}(D_1) \lhd \cdots \lhd \pi^{-1}(D_s) = G.$$
(Here we make essential use of the Fourth Isomorphism Theorem, in particular, $\pi^{-1}(D_i)/\pi^{-1}(D_{i-1}) \cong D_i/D_{i-1}$ for $i = 1, \ldots, s$.)

The proof of (2) will be given in the next lecture. □

Definition 3.4.4. A group G is called **solvable** if there exists a chain of subgroups
$$\{1\} = G_0 \lhd G_1 \lhd \cdots \lhd G_s = G$$
such that G_i/G_{i-1} is abelian for $i = 1, \ldots, s$.

Corollary 3.4.5. *For a finite group G, G is solvable if and only if all of the composition factors of G are of the form $\mathbf{Z}_p$.*

Example 3.4.6. The group of upper triangular invertible matrices is solvable.
$$G = \left\{ \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix} \in \mathrm{GL}_3(\mathbb{C}) \right\} \supseteq N = \left\{ \begin{pmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{GL}_3(\mathbb{C}) \right\} \supseteq N' = \left\{ \begin{pmatrix} 1 & 0 & * \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{GL}_3(\mathbb{C}) \right\}.$$
The subquotients are
$$G/N \cong (\mathbb{C}^\times, \cdot)^3, \quad N/N \cong (\mathbb{C}, +)^2, \quad N' \cong (\mathbb{C}, +).$$

An interesting deep theorem for solvable group is the following.

Theorem 3.4.7 (Philip Hall)**.** *The finite group G is solvable if and only if for every divisor n of $|G|$ such that $\gcd\big(n, \frac{|G|}{n}\big) = 1$, G has a subgroup of order n.*

4. Jordan–Hölder theorem, simplicity of A_n, and direct product groups

4.1. **Jordan–Hölder theorem.**

Theorem 4.1.1 (Jordan–Hölder)**.** *Assume that a group G has the following two composition series*

$$\{1\} = A_0 \lhd A_1 \lhd \cdots \lhd A_m = G \quad \textit{and} \quad \{1\} = B_0 \lhd B_1 \lhd \cdots \lhd B_n = G,$$

then $m = n$, and there exists a bijection $\sigma : \{1, \dots, m\} \to \{1, \dots, n = m\}$ such that

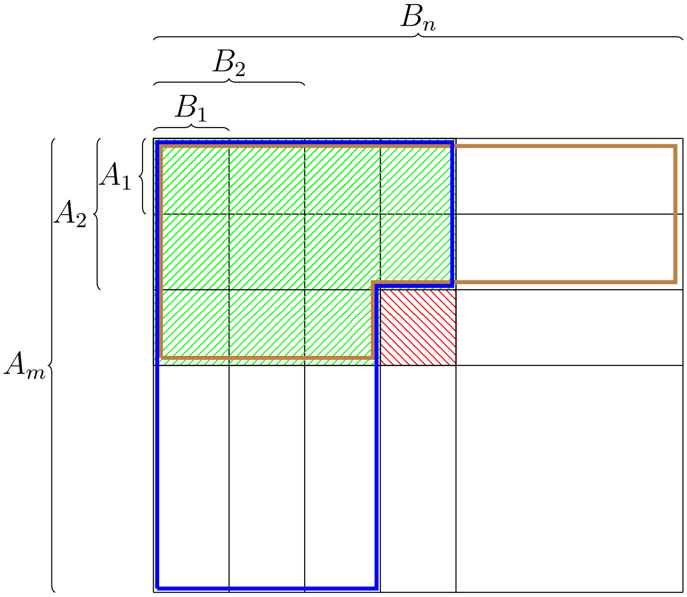
$$A_{\sigma(i)}/A_{\sigma(i)-1} \simeq B_i/B_{i-1}.$$

4.1.2. *Toy model.* A set theoretic version.

Let X be a set with two filtrations.

$$\emptyset = A_0 \subseteq A_1 \subseteq \cdots \subseteq A_m = X, \qquad \emptyset = B_0 \subseteq B_1 \subseteq \cdots \subseteq B_n = X.$$

We use the following picture to explain the situation.



Then we must have for every i, j,

(4.1.2.1)

$$\big(A_{i-1} \cup (A_i \cap B_j)\big) \big\backslash \big(A_{i-1} \cup (A_i \cap B_{j-1})\big) \;=\; \big(B_{j-1} \cup (A_i \cap B_j)\big) \big\backslash \big(B_{j-1} \cup (A_{i-1} \cap B_j)\big).$$

Here $B_{j-1} \cup (A_{i-1} \cap B_j)$ is the blue part and $A_{i-1} \cup (A_i \cap B_{j-1})$ is the brown part. The equality can be seen as both parts represent the shaded red area.

To make the proof a bit more effective, we can first show that both sides are the same as

$$\big(A_i \cap B_j\big) \big\backslash \big((A_i \cap B_{j-1}) \cup (A_{i-1} \cap B_j)\big),$$

where the latter set is the green shaded area. (Indeed, to identify the above complement with the left hand side of (4.1.2.1), we may intersect both terms with B_j; and to identify the above complement with the left hand side of (4.1.2.1), we may intersect both terms with A_i.)

The above proof is of course trivial, but we will see quickly how that help us understand the proof of Jordan–Hölder theorem.

4.1.3. *Proof of Theorem 4.1.1.* We prove a slightly stronger version: let G be a group. Suppose that we are given two chains of subgroups

$$\{1\} = A_0 \trianglelefteq A_1 \trianglelefteq \cdots \trianglelefteq A_m = G, \qquad \{1\} = B_0 \trianglelefteq B_1 \trianglelefteq \cdots \trianglelefteq B_n = G.$$

Then we have

(1) $A_{i-1}(A_i \cap B_{j-1})$ is a normal subgroup of the group $A_{i-1}(A_i \cap B_j)$;
(2) $(A_{i-1} \cap B_j)B_{j-1}$ is a normal subgroup of the group $(A_i \cap B_j)B_{j-1}$;
(3) and we have an isomorphism

$$\frac{A_{i-1}(A_i \cap B_j)}{A_{i-1}(A_i \cap B_{j-1})} \cong \frac{(A_i \cap B_j)B_{j-1}}{(A_{i-1} \cap B_j)B_{j-1}}. \tag{4.1.3.1}$$

This in particular shows that one may refine both chains of subgroups into (setting $A'_{ij} = A_{i-1}(A_i \cap B_j)$ and $B'_{ij} = (A_i \cap B_j)B_{j-1}$)

$$\{1\} = A_0 = A'_{10} \trianglelefteq A'_{11} \trianglelefteq \cdots \trianglelefteq A'_{1n} = A_1 = A'_{20} \trianglelefteq \cdots \trianglelefteq A'_{2n} = A_2 = A'_{30} \trianglelefteq \cdots \trianglelefteq A'_{m-1,n} = A_n = G,$$

$$\{1\} = B_0 = B'_{01} \trianglelefteq B'_{11} \trianglelefteq \cdots \trianglelefteq B'_{m1} = B_1 = B'_{02} \trianglelefteq \cdots \trianglelefteq B'_{m2} = B_2 = B'_{03} \trianglelefteq \cdots \trianglelefteq B'_{m,n-1} = B_n = G,$$

so that $A'_{ij}/A'_{i,j-1} \cong B'_{ij}/B'_{i-1,j}$.

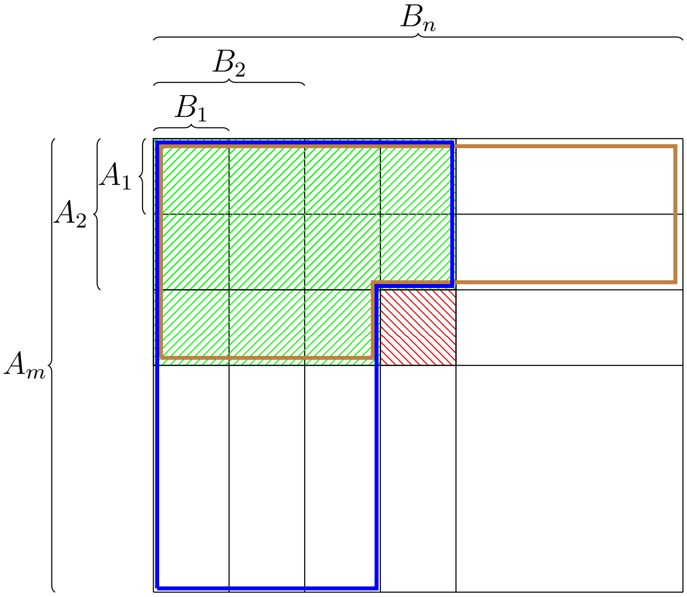
This means that in the special case of the theorem when A_i/A_{i-1} and B_j/B_{j-1} are simple groups, $A_i/A_{i-1} \cong A'_{i,\sigma(i)}/A'_{i,\sigma(i)-1}$ for a unique $\sigma(i) \in \{1, \dots, n\}$ and $B_j/B_{j-1} \cong B'_{\sigma(j),j}/B'_{\sigma(j)-1,j}$ for a unique $\tau(j) \in \{1, \dots, m\}$. It is clear that $m = n$, and σ and τ are inverse of each other. Moreover, (4.1.3.1) implies that $A_i/A_{i-1} \cong B_{\sigma(i)}/B_{\sigma(i)-1}$.

Now we return to prove the stronger version of the theorem above. We first check (1) and (2). By symmetry, it suffices to prove (1). We first show that $A_{i-1}(A_i \cap B_j)$ is a subgroup of G. Indeed, viewing both A_{i-1} and $A_i \cap B_j$ as a subgroup of A_i, A_{i-1} is normal; so $A_{i-1}(A_i \cap B_j)$ is a subgroup of A_i (and hence of G).

Next, we observe that $B_{j-1} \trianglelefteq B_j$ implies that $(A_i \cap B_{j-1}) \trianglelefteq (A_i \cap B_j)$. To show that $A_{i-1}(A_i \cap B_{j-1}) \trianglelefteq A_{i-1}(A_i \cap B_j)$, take $a \in A_{i-1}$, $b \in A_i \cap B_{j-1}$, $\alpha \in A_{i-1}$, and $\beta \in A_i \cap B_j$, we have

$$(\alpha\beta)(ab)(\alpha\beta)^{-1} = \alpha\beta ab\beta^{-1}\alpha^{-1} = \alpha \cdot \underbrace{\beta a \beta^{-1}}_{\text{in } A_{i-1}} \cdot \underbrace{\beta b \beta^{-1}}_{\text{in } A_i \cap B_j} \alpha^{-1} \in A_{i-1}(A_i \cap B_j)A_{i-1} = A_{i-1}(A_i \cap B_j).$$

(The last equality uses that $A_{i-1}(A_i \cap B_j)$ is a group.) This completes the proof of (1), and the two quotients in (4.1.3.1) makes sense.



As suggested by the proof in the toy model, we hope to prove the following:

$$\frac{A_{i-1}(A_i \cap B_j)}{A_{i-1}(A_i \cap B_{j-1})} \cong \frac{A_i \cap B_j}{(A_{i-1} \cap B_j) \cdot (A_i \cap B_{j-1})} \cong \frac{(A_i \cap B_j)B_{j-1}}{(A_{i-1} \cap B_j)B_{j-1}}. \tag{4.1.3.2}$$

By symmetry, it suffices to prove the left isomorphism. We construct a homomorphism

$$\begin{tikzcd} \phi : A_i \cap B_j \arrow[r] & A_{i-1}(A_i \cap B_j) \arrow[r, two heads] & \dfrac{A_{i-1}(A_i \cap B_j)}{A_{i-1}(A_i \cap B_{j-1})} \\ a \arrow[r, mapsto] & a \arrow[r, mapsto] & aA_{i-1}(A_i \cap B_{j-1}). \end{tikzcd}$$

Such homomorphism is clearly surjective. It suffices to find its kernel.

$$\ker \phi = (A_i \cap B_j) \cap \big(A_{i-1}(A_i \cap B_{j-1})\big).$$

Let $a \in A_{i-1}$ and $\beta \in A_i \cap B_{j-1}$. Then

$$a\beta \in B_j \quad \Rightarrow \quad a \in B_j \cdot \beta^{-1} = B_j.$$

So $\ker \phi \subseteq (A_{i-1} \cap B_j)(A_i \cap B_{j-1})$.

Conversely, we clearly have

$$(A_{i-1} \cap B_j)(A_i \cap B_{j-1}) \subseteq A_i \cap B_j \cap \big(A_{i-1}(A_i \cap B_{j-1})\big).$$

Thus, we have $\ker \phi = (A_{i-1} \cap B_j)(A_i \cap B_{j-1})$. By the first isomorphism theorem, we deduce

$$\frac{A_i \cap B_j}{(A_{i-1} \cap B_j) \cdot (A_i \cap B_{j-1})} \cong \frac{A_{i-1}(A_i \cap B_j)}{A_{i-1}(A_i \cap B_{j-1})}.$$

This completes the proof of (3), and the Jordan–Hölder theorem. □

Remark 4.1.4. In fact, what we proved is Zassenhaus Lemma. Let H and K be subgroups of a group G and let H^* and K^* be normal subgroups of H and K, respectively. Then

(1) $H^*(H \cap K^*)$ is a normal subgroup of $H^*(H \cap K)$.
(2) $K^*(H^* \cap K)$ is a normal subgroup of $K^*(H \cap K)$.

(3) $\dfrac{H^*(H\cap K)}{H^*(H\cap K^*)} \cong \dfrac{H\cap K}{(H^*\cap K)(H\cap K^*)} \cong \dfrac{K^*(H\cap K)}{K^*(H^*\cap K)}$.

4.2. **Alternating groups.** One important example of composition series is ($n \geq 5$)

$$\{1\} \leq A_n \overset{2}{\trianglelefteq} S_n.$$

Definition 4.2.1. In the permutation group S_n, recall that for distinct numbers $a_1, \ldots, a_m \in \{1, \ldots, n\}$, one has an m**-cycle** $\sigma = (a_1 a_2 \cdots a_m)$:

$$\begin{tikzcd} a_1 \arrow[r, mapsto, "\sigma"] & a_2 \arrow[r, mapsto, "\sigma"] & \cdots \arrow[r, mapsto, "\sigma"] & a_m \arrow[lll, mapsto, bend left, "\sigma"] \end{tikzcd}$$

A 2-cycle (xy), for $x, y \in \{1, \ldots, n\}$ distinct, is called a **transposition**.

Remark 4.2.2. As $(a_1 \ldots a_m) = (a_1 a_m)(a_1 a_{m-1}) \cdots (a_1 a_2)$, every element of S_n is a product of transpositions.

Properties 4.2.3. Before proceeding, we point out a key observation: for $\sigma \in S_n$, we have

$$\sigma(a_1, \ldots, a_m)\sigma^{-1} = (\sigma(a_1), \ldots, \sigma(a_m)). \tag{4.2.3.1}$$

This can be proved by noting:

$$\begin{tikzcd} \sigma(a_i) \arrow[r, mapsto, "\sigma^{-1}"] & a_i \arrow[r, mapsto, "{(a_1,\ldots,a_m)}"] & a_{i+1} \arrow[r, mapsto, "\sigma"] & \sigma(a_{i+1}). \end{tikzcd}$$

Definition 4.2.4. Define the following for $\sigma \in S_n$

$$\Delta := \prod_{1\leq i<j\leq n} (x_i - x_j), \qquad \sigma(\Delta) := \prod_{1\leq i<j\leq n} (x_{\sigma(i)} - x_{\sigma(j)}) \in \{\pm\Delta\}.$$

For each $\sigma \in S_n$, define $\mathrm{sgn}(\sigma) \in \{\pm 1\}$ so that $\sigma(\Delta) = \mathrm{sgn}(\sigma)\Delta$.

We call $\mathrm{sgn}(\sigma)$ the **sign** of σ.

Proposition 4.2.5. *The map* $\mathrm{sgn} : S_n \to \{\pm 1\}$ *is a homomorphism.*

Proof. By definition, for $\sigma, \tau \in S_n$, we have

$$\mathrm{sgn}(\sigma\tau) = \frac{\prod_{1\leq i<j\leq n} (x_{\sigma\tau(i)} - x_{\sigma\tau(j)})}{\prod_{1\leq i<j\leq n} (x_i - x_j)}.$$

$$\begin{tikzcd} \mathrm{sgn}(\sigma)\cdot\mathrm{sgn}(\tau) = & \boxed{\displaystyle\prod_{1\leq i'<j'\leq n} \frac{(x_{\sigma(i')} - x_{\sigma(j')})}{(x_{i'} - x_{j'})}} \arrow[d, equal, "{i'=\tau(i), j'=\tau(j)}"] & \cdot \ \dfrac{\prod_{1\leq i<j\leq n} (x_{\tau(i)} - x_{\tau(j)})}{\prod_{1\leq i<j\leq n} (x_i - x_j)} \\ & \displaystyle\prod_{1\leq i<j\leq n} \frac{(x_{\sigma\tau(i)} - x_{\sigma\tau(j)})}{(x_{\tau(i)} - x_{\tau(j)})} & \end{tikzcd}$$

Here the vertical equality holds (or rather we are allowed to make the substitution $i' = \tau(i)$ and $j' = \tau(j)$) because in the product of $\dfrac{(x_{\sigma(i')} - x_{\sigma(j')})}{(x_{i'} - x_{j'})}$, we can swap the order of i' and j' (and thus release the constraint that $i' < j'$).

From above, we may cancel the terms $\prod\limits_{1\leq i<j\leq n} \big(x_{\tau(i)} - x_{\tau(j)}\big)$ and thus get $\mathrm{sgn}(\sigma\tau) = \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau)$. □

Definition 4.2.6. The *normal* subgroup $A_n := \ker\big(\mathrm{sgn} : S_n \to \{\pm 1\}\big)$ is called the **alternating group**.

Properties 4.2.7. (1) $A_n \lhd S_n$ and $S_n/A_n \cong \{\pm 1\}$. In particular,

$$|A_n| = |S_n| \,/\, |\{\pm 1\}| = \frac{n!}{2}.$$

(2) We claim that sgn(transpotion) $= -1$. Indeed, $\mathrm{sgn}((12)) = -1$ because for $\sigma = (12)$,

$$\sigma(\Delta) = \prod_{1\leq i<j\leq n} (x_{\sigma(i)} - x_{\sigma(j)}) = (x_2 - x_1) \prod_{\substack{1\leq i<j\leq n \\ j\geq 3}} (x_i - x_j) = -\Delta.$$

For a general transposition (ab), fix $\tau \in S_n$, such that $\tau(1) = a$ and $\tau(2) = b$. Then (4.2.3.1) implies that $\tau(12)\tau^{-1} = (ab)$. Thus,

$$\mathrm{sgn}((ab)) = \mathrm{sgn}(\tau) \cdot \mathrm{sgn}((12)) \cdot \mathrm{sgn}(\tau)^{-1} = \mathrm{sgn}((12)) = -1.$$

Definition 4.2.8. From the discussion above, we see that for $\sigma \in S_n$,

$$\mathrm{sgn}(\sigma) = (-1)^{\text{number of transpositions in the factorization of } \sigma}.$$

So we call such σ

- an **even permutation** if $\mathrm{sgn}(\sigma) = 1$,
- an **odd permutation** if $\mathrm{sgn}(\sigma) = -1$.

In particular, $A_n = \{\sigma \in S_n \,|\, \sigma \text{ is an even permutation}\}$.

Theorem 4.2.9. *When $n \geq 5$, A_n is a simple group.*

Remark 4.2.10. (1) $A_3 = \langle(123)\rangle$ is a cyclic group of order 3.

(2) $A_4 \trianglerighteq \big\{1, (12)(34), (14)(23), (13)(24)\big\} \cong \mathbf{Z}_2^2$.

(3) It is known that a simple group of order 60 is isomorphic to A_5. (It is the smallest non-commutative simple group.)

Proof of Theorem 4.2.9. Recall that a 3-cycle (ijk) always belong to A_n. We will prove three statements, together they prove Theorem 4.2.9.

(1) A_n is generated by all 3-cycles (true with $n \geq 3$).

Indeed, $(a,b)(c,d) = (a,c,b)(a,c,d)$ and $(a,c)(a,b) = (a,b,c)$.

(2) If a normal subgroup $N \trianglelefteq A_n$ contains a 3-cycle, then it contains all 3-cycles (true for $n \geq 3$).

Indeed, Assume that N contains the 3-cycle (i,j,k). Note that, for every $\sigma \in S_n$, either $\sigma \in A_n$ or $\sigma(i,j) \in A_n$. Then (4.2.3.1) implies that

- either $\sigma(i,j,k)\sigma^{-1} = (\sigma(i), \sigma(j), \sigma(k)) \in N$, or
- $\sigma(i,j)(i,j,k)(\sigma(i,j))^{-1} = \sigma(j,i,k)\sigma^{-1} = (\sigma(j), \sigma(i), \sigma(k)) \in N$ (but then we have $(\sigma(j), \sigma(i), \sigma(k))^2 = (\sigma(i), \sigma(j), \sigma(k)) \in N$).

So N always contains $(\sigma(i), \sigma(j), \sigma(k))$ for every $\sigma \in S_n$, and thus N contains all 3-cycles.

(3) If $\{1\} \neq N \lhd A_n$ is a nontrivial normal subgroup, then N contains all 3-cycles.

Take a nontrivial element $\sigma \in N$. We separate several cases:

(a) If σ is a product of disjoint cycles, at least one cycle has length > 3, i.e. $\sigma = \mu(a_1, a_2, \dots, a_r)$ with $r > 3$, then we have

$$(a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1} = \mu(a_2, a_3, a_1, a_4, a_5, \dots, a_r) \in N.$$

So $\sigma^{-1} \circ (a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1}$ sends

$$\begin{tikzcd}
a_1 \arrow[d, mapsto] & a_2 \arrow[d, mapsto] & a_3 \arrow[d, mapsto] & a_4 \arrow[d, mapsto] & \cdots & a_{r-1} \arrow[d, mapsto] & a_r \arrow[d, mapsto] \\
a_4 \arrow[d, mapsto] & a_3 \arrow[d, mapsto] & a_1 \arrow[d, mapsto] & a_5 \arrow[d, mapsto] & \cdots & a_r \arrow[d, mapsto] & a_2 \arrow[d, mapsto] \\
a_3 & a_2 & a_r & a_4 & \cdots & a_{r-1} & a_1.
\end{tikzcd}$$

It is equal to (a_1, a_3, a_r), a 3-cycle.

(b) Suppose that (a) does not hold, and then σ is a product of disjoint 3-cycles and 2-cycles. It then follows that σ^3 is a product of disjoint 2-cycles and σ^2 is a product of disjoint 3-cycles (and they cannot be both 1). So (by considering σ^3 or σ^2 instead of σ, we are reduced to the case when σ is purely a product of disjoint 3-cycles or a product of disjoint 2-cycles.

(c) If σ is a product of one 3-cycles, we are already done. If σ is a product of more than one disjoint 3-cycles, we write $\sigma = \mu(a_4, a_5, a_6)(a_1, a_2, a_3)$. Then

$$(a_1, a_2, a_4)\sigma(a_1, a_2, a_4)^{-1}\sigma^{-1} \in N,$$

and we compute it as:

$$\begin{tikzcd}
a_1 \arrow[d, mapsto] & a_2 \arrow[d, mapsto] & a_3 \arrow[d, mapsto] & a_4 \arrow[d, mapsto] & a_5 \arrow[d, mapsto] & a_6 \arrow[d, mapsto, "\sigma^{-1}"] \\
a_3 \arrow[d, mapsto] & a_1 \arrow[d, mapsto] & a_2 \arrow[d, mapsto] & a_6 \arrow[d, mapsto] & a_4 \arrow[d, mapsto] & a_5 \arrow[d, mapsto, "\substack{(a_1,a_2,a_4)\sigma(a_1,a_2,a_4)^{-1} \\ =\mu(a_2,a_4,a_3)(a_1,a_5,a_6)}"] \\
a_2 & a_5 & a_4 & a_1 & a_3 & a_6
\end{tikzcd}$$

This is $(a_1, a_2, a_5, a_3, a_4)$, a 5-cycle, and we are reduced to case (a).

(d) If σ is a product of (necessarily even number) of disjoint transpositions, we write $\sigma = \mu(a_1, a_2)(a_3, a_4)$. Then

$$(a_1, a_2, a_3)\sigma(a_1, a_2, a_3)^{-1}\sigma^{-1} = (a_1, a_3)(a_2, a_4) \in N.$$

(This step "removes" the extra transpositions μ.) Write $\sigma' := (a_1, a_3)(a_2, a_4)$. After this, we use the condition $n \geq 5$ to take another number $a_5 \in \{1, \dots, n\}$. Explicit computation shows again that

$$(a_1, a_2, a_5)\sigma(a_1, a_2, a_5)^{-1}\sigma^{-1} = (a_1, a_2, a_5, a_4, a_3) \in N,$$

producing a 5-cycle and hence reduces to case (a).

□

4.3. **Direct products.**

Definition 4.3.1. Let I be an index set and let G_i (for $i \in I$) be a group with operator $\star_i$ and identity e_i.

Define the **direct product** of $(G_i)_{i\in I}$, denoted by $\prod_{i\in I} G_i$ (or $G_1 \times G_2 \times \cdots \times G_n$ if $I = \{1, 2, \ldots, n\}$), to be the group with underlying set $G = \prod_{i\in I} G_i$, with operation

$$(g_i)_{i\in I} \star (h_i)_{i\in I} := (g_i \star_i h_i)_{i\in I}.$$

The identity element is $\{e_i\}_{i\in I}$ and the inverse of $(g_i)_{i\in I}$ is $(g_i^{-1})_{i\in I}$.

For each $j \in I$, there is a natural embedding (= injective homomorphism)

$$\begin{tikzcd} G_j \arrow[r, hook] & G = \prod_{i\in I} G_i \\ g_j \arrow[r, mapsto] & (1, \ldots, 1, g_j, 1, \ldots, 1) \end{tikzcd}$$

$$\uparrow$$

$$j^{\text{th}} \text{ place}$$

This realizes G_j as a *normal* subgroup of $\prod_{i\in I} G_i$ and we have

$$\Big(\prod_{i\in I} G_i\Big)\Big/ G_j \cong \prod_{i\in I\setminus\{j\}} G_i.$$

"Dually", there is a natural projection (= surjective homomorphism)

$$\begin{tikzcd} \pi_j : G \arrow[r, two heads] & G_j \\ (g_i)_{i\in I} \arrow[r, mapsto] & g_j \end{tikzcd}$$

We have $\ker \pi_j \cong \prod_{i\in I\setminus\{j\}} G_i$.

Finally, when G_i's are all isomorphic to a group H and $i = \{1, \ldots, r\}$, we write H^r for $\prod_{i\in I} G_i$.

4.4. **Finitely generated abelian groups.**

Recall that a group G is finitely generated if there exist a finite subset A of G such that $G = \langle A\rangle$.

Theorem 4.4.1 (Fundamental theorem of finitely generated abelian groups)**.** *Let G be a finitely generated abelian group. Then*

$$G \simeq \mathbb{Z}^r \times Z_{n_1} \times Z_{n_2} \times \cdots \times Z_{n_s}$$

for some integers $r \geq 0$, $2 \leq n_1 \leq n_2 \leq \cdots \leq n_s$ satisfying $n_i | n_{i+1}$. Moreover these integers $r, n_1, \ldots, n_s$ are unique.

The integer r is called the **rank** *of the abelian group G.*

We will explain later in the semester that abelian groups = $\mathbb{Z}$-modules. So this theorem will follow from the classification of modules over a PID (Theorem 13.3.4).

The goal in this lecture is to see how to characterize finitely generated abelian groups.

Lemma 4.4.2. *If $m, n \in \mathbb{N}_{\geq 2}$ satisfying* $\gcd(m, n) = 1$*, then*

$$\mathbf{Z}_{mn} \cong \mathbf{Z}_m \times \mathbf{Z}_n.$$

Proof. Consider the group homomorphism

$$\begin{aligned} \mathbf{Z}_{mn} &\xrightarrow{\quad\phi\quad} \mathbf{Z}_m \times \mathbf{Z}_n \\ a &\longmapsto \big(a \bmod m, a \bmod n\big). \end{aligned}$$

We compute the kernel of ϕ:

$$\ker\phi = \left\{ a \bmod mn \,\middle|\, \begin{array}{c} a \equiv 0 \bmod m \\ a \equiv 0 \bmod n \end{array} \right\} = \{0 \bmod mn\}.$$

So ϕ is injective. But $|\mathbf{Z}_{mn}| = |\mathbf{Z}_m|\cdot|\mathbf{Z}_n|$. So ϕ must be a bijection and hence an isomorphism. □

Corollary 4.4.3. *Every finitely generated abelian group is of the form*

$$G = \mathbb{Z}^r \times \big(\mathbf{Z}_{p_1^{r_{11}}} \times \cdots \mathbf{Z}_{p_1^{r_{1s_1}}}\big) \times \big(\mathbf{Z}_{p_2^{r_{21}}} \times \cdots \mathbf{Z}_{p_2^{r_{2s_1}}}\big) \times \cdots$$

Here, r, p_1, p_2, ..., $r_{11}, \dots, r_{1s_1}, r_{21}, \dots$ are unique up to permutation.

Example 4.4.4. Determine whether $\mathbf{Z}_{30} \times \mathbf{Z}_{100}$ is isomorphic to $\mathbf{Z}_{60} \times \mathbf{Z}_{50}$. We write

$$\begin{aligned} \mathbf{Z}_{30} \times \mathbf{Z}_{100} &\simeq \mathbf{Z}_2 \times \mathbf{Z}_3 \times \mathbf{Z}_5 \times \mathbf{Z}_4 \times \mathbf{Z}_{25}, \\ \mathbf{Z}_{60} \times \mathbf{Z}_{50} &\simeq \mathbf{Z}_3 \times \mathbf{Z}_4 \times \mathbf{Z}_5 \times \mathbf{Z}_2 \times \mathbf{Z}_{25}. \end{aligned}$$

So they are isomorphic.

Example 4.4.5. List all abelian groups of order $72 = 8 \times 9$. We list them using the following table:

	$\mathbf{Z}_2^3$	$\mathbf{Z}_2 \times \mathbf{Z}_4$	$\mathbf{Z}_8$
$\mathbf{Z}_3^2$	$\mathbf{Z}_3^2 \times \mathbf{Z}_2^3$	$\mathbf{Z}_3^2 \times \mathbf{Z}_2 \times \mathbf{Z}_4$	$\mathbf{Z}_3^2 \times \mathbf{Z}_8$
$\mathbf{Z}_9$	$\mathbf{Z}_9 \times \mathbf{Z}_2^3$	$\mathbf{Z}_9 \times \mathbf{Z}_2 \times \mathbf{Z}_4$	$\mathbf{Z}_9 \times \mathbf{Z}_8$

(We explain why $\mathbf{Z}_2 \times \mathbf{Z}_4$ is not isomorphic to $\mathbf{Z}_8$: for every element $(a,b) \in \mathbf{Z}_2 \times \mathbf{Z}_4$, then $4 \cdot (a,b) = 0$; yet not every element in $\mathbf{Z}_8$ is killed by 4.)

Remark 4.4.6. We explain a general method to determine the factors at p. Suppose that G is an abelian group of order p^n. We would like to determine the numbers $r_1, \dots, r_t$ such that

$$G = \mathbf{Z}_{p^{r_1}} \times \cdots \times \mathbf{Z}_{p^{r_t}}.$$

Consider the number of elements in G killed by p:

$$G[p] = \big\{x \,\big|\, p \cdot x = 0\big\} = \big\{(x_1, \dots, x_t) \,\big|\, x_i \text{ is divisible by } p^{r_i - 1}\big\}.$$

Then $\big|G[p]\big| = p^t$. Next we consider $G[p^2]$:

$$G[p^2] = \big\{x \,\big|\, p^2 \cdot x = 0\big\} = \left\{ (x_1, \dots, x_t) \,\middle|\, \begin{array}{ll} x_i \text{ is divisible by } p^{r_i-2} & r_i \geq 2 \\ x_i \text{ arbitrary} & r_i = 1 \end{array} \right\}.$$

Then we have

$$\big|G[p^2]\big| = p^{2|\{i|r_i \geq 2\}| + |\{i|r_i = 1\}|} = p^{2t - |\{i|r_i=1\}|}.$$

Similarly, we can deduce that

$$\big|G[p^3]\big| = p^{3|\{i|r_i \geq 3\}| + 2|\{i|r_i=2\}| + |\{i|r_i=1\}|} = p^{3t - 2|\{i|r_i=1\}| - |\{i|r_i=2\}|}.$$

Continue this, we may recover the numbers r_i from $\big|G[p^n]\big|$'s.

5. Recognizing direct produces, group actions, semi-direct products

5.1. **Recognizing direct products.** We prove a theorem that "recognize" when a group is a direct product (of its subgroups).

Theorem 5.1.1 (Criterion of direct product group)**.** *Suppose that G is a group with subgroups H and K such that*

1. *H and K are normal subgroups of G, and*
2. *$H \cap K = \{1\}$.*

Then $HK \cong H \times K$ (as groups).

Proof. Since both H and K are normal subgroups of G, HK is a normal subgroup of G. Consider the natural map

$$\begin{aligned} \phi : H \times K &\longrightarrow HK \\ (h,k) &\longmapsto hk \end{aligned}$$

- ϕ is a homomorphism. For this, we need to check that, for $h_1, h_2 \in H$ and $k_1, k_2 \in K$, we have

  $$\begin{tikzcd} \phi\big((h_1,k_1)(h_2,k_2)\big) \arrow[r, equal] & \phi(h_1h_2, k_1k_2) \arrow[r, equal] & h_1h_2k_1k_2 \\ \phi(h_1,k_1)\phi(h_2,k_2) \arrow[u, equal, "?"] \arrow[rr, equal] & & h_1k_1h_2k_2. \end{tikzcd}$$

  It suffices to show that

  $$h_2k_1 = k_1h_2, \quad \text{or equivalently} \quad k_1h_2k_1^{-1}h_2^{-1} = 1.$$

  But the normality of K and H implies that

  $$\underbrace{k_1h_2k_1^{-1}}_{\text{in } H}h_2^{-1} \in H \quad \text{and} \quad k_1\underbrace{h_2k_1^{-1}h_2^{-1}}_{\text{in } K} \in K.$$

  So $k_1h_2k_1^{-1}h_2^{-1} \in H \cap K = \{1\}$.
- ϕ is clearly surjective.
- $\ker\phi = \big\{(k,h) \in K \times H \,\big|\, kh = 1\big\}$. But this condition implies that $k = h^{-1} \in K \cap H = \{1\}$. So $\ker\phi = \{1\}$.

Summing up above, we deduce that ϕ is an isomorphism. □

5.2. **Concept of group actions.** Next, we will discuss a very important concept: group actions. The basic example is S_n "permuting" numbers in $\{1, 2, \ldots, n\}$. We say that S_n *acts on the set* $\{1, 2, \ldots, n\}$. We hope to generalize this notion of a group "acting" or "permuting" elements in another set.

Definition 5.2.1. Let G be a group and X a set. A **left G-action on X** is a map

$$\begin{aligned} G \times X &\longrightarrow X \\ (g,x) &\longmapsto g \cdot x \end{aligned}$$

satisfying the following conditions:

1. for any $x \in X$, $e \cdot x = x$,

(2) for any $g, h \in G$ and $x \in X$, we have

$$g \cdot (h \cdot x) = (gh) \cdot x.$$

Sometimes, we write $G \circlearrowright X$.

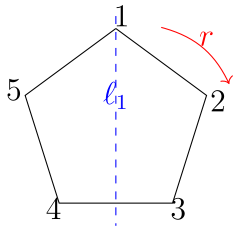
Later, unless otherwise specified, we will always consider left group actions *as opposed to right group actions defined later.*

Remark 5.2.2. The condition implies that for any $g \in G$, $X \to X$ given by $x \mapsto g \cdot x$ is a bijection (because the inverse is given by $x \mapsto g^{-1}x$ as $x \xmapsto{g} g \cdot x \xmapsto{g^{-1}} g^{-1} \cdot g \cdot x = x$.

Example 5.2.3. (1) S_n acts on $X = \{1, 2, \dots, n\}$; we verify Definition 5.2.1(2) as $\sigma(\tau(i)) = (\sigma \cdot \tau)(i)$ for each $i \in X$.

(2) D_{2n} acts on a regular n-gon, by taking the symmetry, where

$$r = \text{rotation clockwise } \frac{2\pi}{n} \quad \text{and} \quad s = \text{reflection about } \ell_1.$$



Symmetry of a pentagon

(3) Let G be a group acting on a set X and H a subgroup of G, then we may *restrict* the G-action on X to an H-action on X.

For example, we may restrict the D_{2n}-action on a regular n-gon to the subgroup $\langle r \rangle$, which will only rotate the regular n-gon.

(4) G-action on itself:

- **left translation action**: for $g \in G$, consider

$$\ell_g : G \longrightarrow G$$
$$\ell_g(x) := gx.$$

This is an action because $\ell_g \circ \ell_h = \ell_{gh}$.

- **right translation action**: for $g \in G$, consider

$$r_g : G \longrightarrow G$$
$$r_g(x) := xg^{-1}.$$

Why do we use g^{-1}? This is because we need $r_g \circ r_h = r_{gh}$ to define a *left action*. We check: for $x \in G$,

$$\begin{array}{ccccc} r_g \circ r_h(x) & = & r_g(xh^{-1}) & = & xh^{-1}g^{-1} \\ \|? & & & & \| \\ r_{gh}(x) & == & x(gh)^{-1} & = & xh^{-1}g^{-1} \end{array}$$

(If we had defined $r'_g(x) = xg$, then $r'_g \circ r'_h(x) = r'_g(xh) = xhg \neq xgh = r'_{gh}(x)$.)

- **conjugation action**: for $g \in G$, consider

$$\mathrm{Ad}_g : G \longrightarrow G$$
$$\mathrm{Ad}_g(x) := gxg^{-1}.$$

This is a "better" action because the map Ad_g is in fact a homomorphism (and also an isomorphism):

$$\mathrm{Ad}_g(x) \cdot \mathrm{Ad}_g(y) = gxg^{-1} \cdot gyg^{-1} = gxyg^{-1} = \mathrm{Ad}_g(xy).$$

There is also a version of right group action.

Definition 5.2.4. Let G be a group and let X be a set. A **right action** of G on X is a map

$$\begin{aligned} X \times G &\longrightarrow X \\ (x, g) &\longmapsto x \cdot g. \end{aligned}$$

such that $x \cdot e = x$ and $(x \cdot g) \cdot h = x \cdot (gh)$ for any $x \in X$ and $g, h \in G$.

For example, right translation by g is a right action:

$$r'_g : G \longrightarrow G$$
$$r'_g(x) := xg.$$

Proposition 5.2.5. *Let G be a group acting on a set X. Then we have a natural homomorphism from G to the permutation group of X:*

$$\begin{aligned} \Phi : G &\longrightarrow S_X \\ g &\longmapsto (\phi_g : x \mapsto g \cdot x). \end{aligned}$$

In fact, given a group action of G on X is equivalent to give a homomorphism $\Phi : G \to S_X$.

Proof. We need to check that $\phi_g \circ \phi_h = \phi_{gh}$ for every $g, h \in G$. Indeed, for $x \in X$,

$$\phi_g \circ \phi_h(x) = \phi_g(h \cdot x) = g \cdot (h \cdot x) = (gh) \cdot x = \phi_{gh}(x).$$

□

Definition 5.2.6. (1) If this homomorphism Φ is injective, we say this action is **faithful**. In this case, we may identify G with a subgroup of S_X.

This is equivalent to require $\ker \Phi = \{1\}$, meaning no nontrivial element of G fixes all elements of X.

(2) If this homomorphism Φ is trivial, i.e. $\phi_g = \mathrm{id}$ for every $g \in G$, we say that the action is **trivial**.

Theorem 5.2.7 (Cayley)**.** *Every group is isomorphic to a subgroup of some symmetry group. If $|G| = n$, then G is isomorphic to a subgroup of S_n.*

Proof. Consider the left translation action; by Proposition 5.2.5, it induces a homomorphism $G \hookrightarrow S_G$. This action is clearly faithful, and thus identify G as a subgroup of S_G. □

Remark 5.2.8. (1) This theorem has historical meaning because groups are first defined as subgroups of S_n. Cayley's theorem says that our abstract definition agrees with the old definition.

(2) Cayley's theorem also suggests: given an abstract group, if we want to understand G on the element level, it is better to let G act on some set X and represent G as a subgroup of S_X. (We will see more examples of this sort in the next lecture.)

5.3. **Automorphism groups.** As we explained earlier, conjugation action is a "better" action of G on itself.

Definition 5.3.1. An **automorphism** of a group G is an isomorphism $\sigma : G \xrightarrow{\simeq} G$. Then

$$\mathrm{Aut}(G) := \big\{\text{automorphisms of } G\big\}$$

forms a group, with

- identity being $\mathrm{id} : G \to G$;
- group action being composition; and
- inverse being the inverse isomorphism.

It is a subgroup of S_G = permutation group of elements of G.

Remark 5.3.2. If we consider the conjugation action of G on itself. It induces a homomorphism

$$\begin{aligned} \mathrm{Ad} : G &\longrightarrow \mathrm{Aut}(G) \subseteq S_G \\ g &\longmapsto (\mathrm{Ad}_g : x \mapsto gxg^{-1}). \end{aligned}$$

Note that we have seen that each Ad_g is a homomorphism.

5.3.3. *Group acting on a group through automorphism.* More generally, we may consider the case when X is also a group, carrying an action of another group G which preserves the group structure on X, namely, for each $g \in G$, the action $\phi_g : X \to X$ is in fact a homomorphism (and hence an isomorphism); or equivalently, the natural homomorphism $\Phi : G \to S_X$ has image in $\mathrm{Aut}(X)$. In this case, we say that G **acts on the group X by automorphisms**.

One may image to generalize this to future situations when X is a set with additional algebraic structure; we may require the G-action on X to preserve the corresponding algebraic structure.

5.4. **Semi-direct products.**

5.4.1. *A prototype of semidirect product.* Consider the situation:

(a) $N \trianglelefteq G$ and $H \leq G$;
(b) $N \cap H = \{1\}$.

Then $NH = \{nh \,|\, n \in N, h \in H\}$ is a subgroup of G.

Note that if H is also a normal subgroup, Theorem 5.1.1 implies that $NH \cong N \times H$ and N and H commutes.

In general, every element of NH can be written uniquely as nh with $n \in N$ and $h \in H$. If we compute the product of n_1h_1 and n_2h_2:

$$n_1h_1 \cdot n_2h_2 = n_1 \underbrace{h_1n_2h_1^{-1}}_{\text{in } N} \cdot h_1h_2.$$

We see that the "H-coordinate" is multiplicative, but not the "N-coordinate".

The following "partially reverses" the above discussion.

Definition 5.4.2. Let N and H be groups, and let $\phi : H \to \mathrm{Aut}(N)$ be a homomorphism. (Then H acts on N preserving the group structure; see § 5.3.3). For $h \in H$, we write $\phi_h = \phi(h) : N \to N$ for the corresponding automorphism.

We define the **semi-direct product** $N \rtimes H = N \rtimes_\phi H$ to be

$$N \rtimes H := \big\{(n,h) \,\big|\, n \in N, h \in H\big\}$$
$$(n_1, h_1) \cdot (n_2, h_2) = (n_1 \cdot \phi_{h_1}(n_2), h_1 h_2).$$

(We see that the H-coordinate is the usual multiplication, but the multiplication in the N-coordinate is "twisted by ϕ".)

We now check the group action is associative and that it has inverses.

$$\begin{tikzcd}
\big((n_1,h_1)(n_2,h_2)\big)(n_3,h_3) \arrow[r, equal, "?"] \arrow[d, equal, "\text{computing first multiplication}"] & (n_1,h_1)\big((n_2,h_2)(n_3,h_3)\big) \arrow[d, equal, "\text{computing second multiplication}"] \\
(n_1\phi_{h_1}(n_2), h_1h_2)(n_3,h_3) \arrow[d, equal, "\text{definition of multiplication}"] & (n_1,h_1)(n_2\phi_{h_2}(n_3), h_2h_3) \arrow[d, equal, "\text{definition of multiplication}"] \\
(n_1\phi_{h_1}(n_2)\phi_{h_1h_2}(n_3), h_1h_2h_3) \arrow[d, equal, "\phi_{h_1h_2}=\phi_{h_1}\circ\phi_{h_2}"] & \big(n_1 \cdot \phi_{h_1}\big(n_2\phi_{h_2}(n_3)\big), h_1h_2h_3\big) \arrow[d, equal, "\phi_{h_1}\text{ is a homomorphism}"] \\
(n_1\phi_{h_1}(n_2) \cdot \phi_{h_1} \circ \phi_{h_2}(n_3), h_1h_2h_3) \arrow[r, equal] & \big(n_1\phi_{h_1}(n_2) \cdot \phi_{h_1}(\phi_{h_2}(n_3)), h_1h_2h_3\big).
\end{tikzcd}$$

$$(n,h)(\phi_{h^{-1}}(n), h^{-1}) = (n \cdot \phi_h(\phi_{h^{-1}}(n)), hh^{-1}) = (n \cdot n^{-1}, 1) = (1,1).$$

The sets $\{(n,1) \,|\, n \in N\} \subseteq N \rtimes H$ and $\{(1,h) \,|\, h \in H\} \subseteq N \rtimes H$ are subgroups. They may be viewed as the groups N and H naturally embedded in $N \rtimes H$.

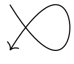
There is a surjective homomorphism

$$\begin{aligned} \pi : N \rtimes H &\longrightarrow H \\ (n,h) &\longrightarrow h \end{aligned}$$

and $\ker \pi$ is the subgroup N; so in particular, N is a normal subgroup

Remark 5.4.3. There are two ways to remember the notation $N \rtimes H$.

(1) In the notation $N \rtimes H$, the triangle part is towards H; so H is the normal subgroup.

(2) Recall that we denote the group action as $G \circlearrowright X$. We may think the symbol $\rtimes$ as a "twisted" way of writing the action: N  H

Convention 5.4.4. Another way to think of the definition of semidirect product is to define $N \rtimes H$ as the set of formal products $\big\{nh \,\big|\, n \in N,\ h \in H\big\}$ subject to the rule that

$$n_1h_1 \cdot n_2h_2 = n_1h_1n_2h_1^{-1}h_1h_2 = (n_1\phi_{h_1}(n_2)) \cdot (h_1h_2).$$

One can see that this agrees with the original definition of $N \rtimes H$.

We can use a different convention to write $H \ltimes N$ instead, to mean exactly the *same* semidirect product. Except now, we may write the elements in $H \ltimes N$ as formal products $\big\{hn \,\big|\, h \in H,\ n \in N\big\}$ subject to the product rule that

$$h_1n_1 \cdot h_2n_2 = h_1h_2 \cdot (h_2^{-1}n_1h_2 \cdot n_2) = h_1h_2 \cdot (\phi_{h_2^{-1}}(n_1)n_2).$$

In other words, if we choose to write the elements in $H \ltimes N$ as pairs (h, n), then the product rule is

$$(h_1, n_1) \cdot (h_2, n_2) = (h_1h_2, \phi_{h_2^{-1}}(n_1)n_2).$$

Proposition 5.4.5 (Recognizing semidirect products)**.** *Let G be a group, and let $N \trianglelefteq G$ a normal subgroup and $H \leq G$ a subgroup. Suppose that $N \cap H = \{1\}$. Then NH is a subgroup of G and $NH \cong N \rtimes H$ is a semidirect product.*

Proof. We have proved that NH is a subgroup of G. As N is a normal subgroup of G, the conjugation action for each $h \in H$ defines an automorphism $\mathrm{Ad}_h : N \to N$ given by $\mathrm{Ad}_h(n) = hnh^{-1}$. Collectively, this defines a homomorphism $\mathrm{Ad} : H \to \mathrm{Aut}(N)$. □

The following proposition is left as an exercise.

Proposition 5.4.6. *Let N and H be groups and let $\phi : H \to \mathrm{Aut}(N)$ be a homomorphism. The following are equivalent (TFAE):*

- *The identity map between $N \rtimes_\phi H$ and $N \times H$ is a group homomorphism (and hence an isomorphism);*
- *ϕ is the trivial homomorphism from $H \to \mathrm{Aut}(N)$;*
- *N is a normal subgroup of $N \rtimes H$.*

Example 5.4.7. A typical example of semi-direct product comes from the following.

Recall that $\mathbf{Z}_n$ is the group of modulo n residual classes. Then $\mathrm{Aut}(\mathbf{Z}_n, +) \cong \mathbf{Z}_n^\times = \big\{a \bmod n \,\big|\, \gcd(a, n) = 1\big\}$. Here, for $a \bmod n \in \mathbf{Z}_n^\times$, the corresponding automorphism of $\mathbf{Z}_n$ is given by

$$\begin{aligned} \phi_a : \mathbf{Z}_n &\longrightarrow \mathbf{Z}_n \\ x &\longmapsto ax \bmod n \end{aligned}$$

This gives rise to a semi-direct product $\mathbf{Z}_n \rtimes \mathbf{Z}_n^\times$.

We can visualize the group $\mathbf{Z}_n \rtimes \mathbf{Z}_n^\times$ as

$$\mathbf{Z}_n \rtimes \mathbf{Z}_n^\times \cong \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \in \mathrm{M}_{2\times 2}(\mathbf{Z}_n) \;\middle|\; a \in \mathbf{Z}_n^\times \right\}.$$

One checks $\begin{pmatrix} a_1 & b_1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_2 & b_2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_1a_2 & a_1b_2 + b_1 \\ 0 & 1 \end{pmatrix}$, comparing to our convention $(b_1, a_1)(b_2, a_2) = (b_1 + a_1b_2, a_1a_2)$. (Note that the normal subgroup is on the left here.)

We can also consider subgroups of $\mathbf{Z}_n^\times$. For example, $\{\pm 1\} \in \mathbf{Z}_n^\times = \mathrm{Aut}(\mathbf{Z}_n)$. We have

$$\begin{aligned} \mathbf{Z}_n \rtimes \{\pm 1\} &\xrightarrow{\cong} D_{2n} \\ (a, 1) &\longmapsto r^a \\ (a, -1) &\longmapsto r^a s. \end{aligned}$$

(Note that the relation $srs = r^{-1}$ in D_{2n} corresponds to semidirect product relation of ϕ-action on $\mathbf{Z}_n$.)

Example 5.4.8. Another class of semidirect product is the block upper-triangular matrices. Take for example three positive integers ℓ, m, n and consider the following groups

$$P = \begin{pmatrix} \mathrm{GL}_\ell(\mathbb{C}) & \mathrm{Mat}_{\ell\times m}(\mathbb{C}) & \mathrm{Mat}_{\ell\times n}(\mathbb{C}) \\ 0 & \mathrm{GL}_m(\mathbb{C}) & \mathrm{Mat}_{m\times n}(\mathbb{C}) \\ 0 & 0 & \mathrm{GL}_n(\mathbb{C}) \end{pmatrix} \quad L = \begin{pmatrix} \mathrm{GL}_\ell(\mathbb{C}) & 0 & 0 \\ 0 & \mathrm{GL}_m(\mathbb{C}) & 0 \\ 0 & 0 & \mathrm{GL}_n(\mathbb{C}) \end{pmatrix}$$

$$U = \begin{pmatrix} I_\ell & \mathrm{Mat}_{\ell\times m}(\mathbb{C}) & \mathrm{Mat}_{\ell\times n}(\mathbb{C}) \\ 0 & I_m & \mathrm{Mat}_{m\times n}(\mathbb{C}) \\ 0 & 0 & I_n \end{pmatrix}$$

Here P is usually called a *parabolic subgroup*, U is called its *unipotent radical*, and L is called the *Levi subgroup* of P. These types of groups can be generalized to the situations of other "algebraic groups", and are very important constructions in algebraic groups and representations.

Taking our concrete example at hand, one can see that U is a normal subgroup of P, but L is not. So we have a semi-direct product

$$P = U \rtimes L.$$

Example 5.4.9. Let p and q be distinct primes such that $p \mid (q-1)$. We may use Example 5.4.7 construct nonabelian groups of order pq which are semidirect products.

It is known (will be proved later) that $\mathbf{Z}_q^\times$ is a cyclic group of order $q-1$. So it must contain a *unique* subgroup of order p. This in particular gives a homomorphism $\mathbf{Z}_p \hookrightarrow \mathbf{Z}_q^\times = \mathrm{Aut}(\mathbf{Z}_q)$ (the homomorphism is not unique but see Fact 5.4.10) By definition of semidirect product, this gives a semidirect product group $\mathbf{Z}_q \rtimes \mathbf{Z}_p$ of order pq.

For example, we describe $\mathbf{Z}_7 \rtimes \mathbf{Z}_3$ as follows. Consider the homomorphisms

$$\begin{aligned} \phi_1, \phi_2 : \mathbf{Z}_3 &\longrightarrow \mathbf{Z}_7^\times \\ \phi_1 : 0, 1, 2 &\longmapsto 1, 2, 4 \\ \phi_2 : 0, 1, 2 &\longmapsto 1, 4, 2. \end{aligned}$$

(The reason that we have these two homomorphisms is that 3 is a primitive element modulo 7, and ϕ_i needs to map 1 to either 3^2 or to 3^4.) Then the corresponding semi-direct product has group structure:

$$(a_1, b_1)(a_2, b_2) = (a_1 + a_2 \cdot 2^{b_1}, b_1 + b_2) \quad \text{or} \quad (a_1 + a_2 \cdot 4^{b_1}, b_1 + b_2)$$

in either cases.

Fact 5.4.10. Let p and q be distinct primes as above.

1. For two different nontrivial homomorphisms $\phi_1, \phi_2 : \mathbf{Z}_p \to \mathbf{Z}_q^\times$, the semi-direct products $\mathbf{Z}_q \rtimes_{\phi_i} \mathbf{Z}_p$ with $i = 1, 2$ are isomorphic.
2. All groups of order pq are either isomorphic to $\mathbf{Z}_{pq}$ (abelian case) or to $\mathbf{Z}_q \rtimes \mathbf{Z}_p$ (nonabelian case).

(This fact is specific to the group $\mathbf{Z}_q \rtimes \mathbf{Z}_p$, and is not true in general.)

We illustrate this using the example above. We can construct an isomorphism

$$\begin{tikzcd} \mathbf{Z}_7 \rtimes_{\phi_1} \mathbf{Z}_3 \arrow[r, "\cong"] & \mathbf{Z}_7 \rtimes_{\phi_2} \mathbf{Z}_3 \\ (a,b) \arrow[r, mapsto] & (a,2b). \end{tikzcd}$$

We need to check that ψ is a homomorphism, namely:

$$\begin{tikzcd} \psi((a,b)(c,d)) \arrow[r, equal, "?"] \arrow[d, equal] & \psi((a,b)) \cdot \psi((c,d)) \arrow[d, equal] \\ \psi(a+2^b \cdot c, b+d) \arrow[d, equal] & (a,2b)(c,2d) \arrow[d, equal] \\ (a+2^b \cdot c, 2b+2d) \arrow[r, equal] & (a+4^{2b}c, 2b+2d). \end{tikzcd}$$

Here the bottom equality uses that $4^{2b} = 16^b \equiv 2^b \bmod 7$.

6. Stablizers, orbits of group actions, class equations

6.1. Stabilizers and orbits of group actions.

Definition 6.1.1. Let G be a group acting on a set X. For each $x \in X$,

- define the **stabilizer subgroup** at x to be $\mathrm{Stab}_G(x) := \{g \in G \,|\, g \cdot x = x\}$; and
- define the **orbit** of x to be $\mathrm{Orb}_G(x) := G \cdot x = \{g \cdot x \,|\, g \in G\} \subseteq X$.

We sometimes write $G\backslash X$ for the set of orbits for the G-action. (In the literature, it might be written as X/G instead. I slightly prefer $G\backslash X$ because the action is from the left.)

Properties 6.1.2. Let G be a group acting on a set X and $x \in X$.

(1) Then $\mathrm{Stab}_G(x)$ is a subgroup of G.

(2) For $x, y \in X$, either $\mathrm{Orb}_G(x) = \mathrm{Orb}_G(y)$ or $\mathrm{Orb}_G(x) \cap \mathrm{Orb}_G(y) = \emptyset$. As a corollary, X may be written as the *disjoint union* of orbits for the G-action:

$$X = \coprod_{\text{orbits } \mathcal{O}} \mathcal{O}.$$

(3) If $y \in \mathrm{Orb}_G(x)$, i.e. $y = g \cdot x$ for some $g \in G$, then $\mathrm{Stab}_G(y) = g\mathrm{Stab}_G(x)g^{-1}$. Namely, the stabilizers at different points of an orbit are conjugate to each other.

Proof. (1) If $g, h \in \mathrm{Stab}_G(x)$, we need to prove that $gh^{-1} \in \mathrm{Stab}_G(x)$, namely $gh^{-1} \cdot x = x$. As $x = hx$, applying h^{-1} to this we deduce that $h^{-1}x = h^{-1}hx = x$. Now applying g to both sides of the equality gives $gh^{-1}x = gx = x$. This verifies that $\mathrm{Stab}_G(x)$ is a subgroup of G.

(2) Suppose that $\mathrm{Orb}_G(x) \cap \mathrm{Orb}_G(y) \neq \emptyset$, say both sets contain z. Then we have $z = g \cdot x = h \cdot y$ for some $g, h \in G$. This implies that for every element $w = kx \in \mathrm{Orb}_G(x)$ ($k \in G$), we have

$$w = kx = kg^{-1}z = kg^{-1}hy \in \mathrm{Orb}_G(y).$$

This proves that $\mathrm{Orb}_G(x) \subseteq \mathrm{Orb}_G(y)$. A symmetric argument shows that $\mathrm{Orb}_G(y) \subseteq \mathrm{Orb}_G(x)$.

(3) This can be shown as follows.

$$\begin{aligned} h \in \mathrm{Stab}_G(y) &\iff hy = y \\ &\iff hgx = gx \\ &\iff g^{-1}hgx = x \\ &\iff g^{-1}hg \in \mathrm{Stab}_G(x) \\ &\iff h \in g\mathrm{Stab}_G(x)g^{-1}. \end{aligned}$$

□

An important particular case of group action is the conjugation action of a group on itself.

Definition 6.1.3. Consider the group G acting on itself by conjugation: for $g \in G$, $\mathrm{Ad}_g : G \to G$ given by $\mathrm{Ad}_g(x) = gxg^{-1}$.

(1) Two elements $a, b \in G$ are called **conjugate** if $a = gbg^{-1}$ for some $g \in G$. In other words, $a \in \mathrm{Orb}_G(b)$ or equivalently $b \in \mathrm{Orb}_G(a)$.

(2) The orbits of G under the conjugation action are called **conjugacy classes**.

Example 6.1.4. (1) If G is abelian, the conjugacy class of an element $a \in G$ is just $\{a\}$.

(2) For $G = \mathrm{GL}_n(\mathbb{C})$, every matrix can be conjugated into a Jordan block. So there is a bijection

$$\big\{\text{Conjugacy classes of } G\big\} \longleftrightarrow \big\{\text{Jordan canonical form (with nonzero eigenvalues up to permutation)}\big\}.$$

(3) $G = S_n$, the conjugacy classes are in one-to-one correspondence with partitions of n into sums of positive integers. More precisely, a partition $n = n_1 + n_2 + \cdots + n_t$ corresponds to the conjugacy class of

$$\tau_{\underline{n}} = (1, 2, \ldots, n_1)(n_1+1, n_1+2, \ldots, n_1+n_2)\cdots(n_1+\cdots+n_{t-1}+1, \ldots, n_1+\cdots+n_t).$$

For each $\sigma \in S_n$, $\sigma\tau_{\underline{n}}\sigma^{-1}$ is the same as

$$(\sigma(1), \ldots, \sigma(n_1))(\sigma(n_1+1), \ldots, \sigma(n_1+n_2))\cdots(\sigma(n_1+\cdots+n_{t-1}+1), \ldots, \sigma(n_1+\cdots+n_t)).$$

Such elements run through all elements of S_n, which is the product of disjoint cycles of length n_1, n_2, ..., n_t.

Definition 6.1.5. Let G be a group, H a subgroup, and $S \subseteq G$ a subset.

(1) The subgroup $C_G(S) := \big\{g \in G \,\big|\, \text{for every } s \in S,\ gsg^{-1} = s\big\}$ is called the **centralizer** of S in G.

(2) The subgroup $Z(G) = \big\{g \in G \,\big|\, \text{for all } h \in G,\ ghg^{-1} = h\big\} = C_G(G)$ is called the **center** of G.

(3) The subgroup $N_G(H) = \big\{g \in G \,\big|\, gHg^{-1} = H\big\}$ is called the **normalizer** of H in G.

Properties 6.1.6. (1) When $S = \{g\}$, we write $C_G(g)$ for $C_G(\{g\})$. If we consider the conjugation of G on G, then $\mathrm{Stab}_G(g) = C_G(g)$. (In particular, this is a subgroup.) Moreover, we have $C_G(S) = \bigcap_{g \in S} C_G(g)$; so it is also a subgroup.

(2) The conjugation action induces a homomorphism $\mathrm{Ad} : G \to \mathrm{Aut}(G)$. Then $Z(G) = \ker(\mathrm{Ad})$. (In particular, $Z(G)$ is a normal subgroup of G.)

(3) A subgroup H of G is normal if and only if $N_G(H) = G$.

(4) If we consider the conjugation action of G on the set $\big\{\text{all subgroups of } G\big\}$, given by

$$\mathrm{Ad}_g : H \mapsto gHg^{-1}.$$

Then $N_G(H) = \mathrm{Stab}_G(H)$; so in particular a subgroup.

(5) $N_G(H)$ contains H as a subgroup, and H is normal in $N_G(H)$.

6.2. Description of orbits of group actions.

Definition 6.2.1. Let G be a group acting on two sets X and Y. We say a map $\phi : X \to Y$ is **G-equivariant** if

$$\text{for all } g \in G,\ x \in X,\ \text{ we have } \phi(g \cdot x) = g \cdot \phi(x).$$

Remark 6.2.2. To better understand the above definition of G-equivariant maps, we note that, in the development of mathematical theory, we often have the following process

Algebraic structure on a set	$\rightsquigarrow$	Maps between sets with algebraic structure that preserves the algebraic structures
Vector spaces	$\rightsquigarrow$	linear maps
Groups	$\rightsquigarrow$	Homomorphisms
Sets with group actions	$\rightsquigarrow$	G-equivariant maps

Definition 6.2.3. Let G be a group acting on a set X. We say that the action is **transitive** if

$$\text{for any } x, y \in X, \text{ there exists } g \in G, \text{ such that } x = gy.$$

Proposition 6.2.4. *If a group G acts transitively on a set X, for every element $x \in X$, put $H := \mathrm{Stab}_G(x)$. Then there is a G-*equivariant *bijection*

$$\begin{aligned} \phi : G/H &\xrightarrow{\cong} X \\ gH &\longmapsto gx. \end{aligned}$$

Proof. First, ϕ is well-defined because if $g_1H = g_2H$, then $g_1 = g_2h$ for some $h \in H$. Thus

$$g_1x = g_2hx = g_2x.$$

Second, ϕ is surjective because the G-action on X is transitive.

Third, ϕ is injective because if $\phi(g_1H) = \phi(g_2H)$ form some $g_1, g_2 \in G$,

$$g_1x = g_2x \implies g_2^{-1}g_1x = x \implies g_2^{-1}g_1 \in \mathrm{Stab}_G(x) = H \implies g_1H = g_2H.$$

Lastly, ϕ is G-equivariant because for $g, g' \in G$,

$$g'\phi(gH) = g'gx = \phi(g'gH).$$

□

Corollary 6.2.5. *Let G be a group acting on a set X. Then $G \cong \coprod_{\text{orbits } \mathcal{O}} \mathcal{O}$ (as sets with G-action).*

For each $x \in X$, G acts transitively on $\mathrm{Orb}_G(x)$, we have

$$\mathrm{Orb}_G(x) = G/\mathrm{Stab}_G(x).$$

Summing up this, we have

$$X \simeq \coprod_{G\text{-orbits } G\cdot x} G/\mathrm{Stab}_G(x).$$

6.3. **Class equations.**

Theorem 6.3.1. *Let G be a finite group (acting on itself by conjugation).*

(1) *For each $g \in G$, the number of elements in its conjugacy class is*

$$\left|\mathrm{Ad}_G(g)\right| = |G|/|C_G(g)| = [G : C_G(g)].$$

(2) *(Class equation) If $g_1, g_2, \dots, g_r$ are representatives of conjugacy classes of G that are not contained in $Z(G)$, then*

$$|G| = |Z(G)| + \sum_{i=1}^{r} \big[G : C_G(g_i)\big]. \tag{6.3.1.1}$$

Moreover, $\big[G : C_G(g_i)\big] \geq 2$ for each i.

Proof. (1) is clear from Proposition 6.2.4: $\mathrm{Ad}_G(g) = G/C_G(g)$.

(2) Consider the conjugation action of G on itself. By Corollary 6.2.5, we have

$$|G| = \sum_{\substack{\text{conjugacy} \\ \text{classes } \mathrm{Ad}_G(x)}} \left|\mathrm{Ad}_G(x)\right|.$$

There are two types of orbits.

- If $\mathrm{Ad}_G(x)$ consists of only one element, namely x, then $gxg^{-1} = x$ for every $g \in G$, i.e. $x \in Z(G)$.
- If $\mathrm{Ad}_G(x)$ consists of more than one element, we just have $\big|\mathrm{Ad}_G(x)\big| = \big[G : C_G(g_i)\big]$.

Combining these two cases proves (6.3.1.1). □

Example 6.3.2. Let $G = S_5$. Then $Z(G) = \{1\}$. The class equation reads

Partition type	representative	stabilizer	size of conjugacy class
$1+1+1+1+1$	(1)	S_5	$\frac{120}{120} = 1$
$1+1+1+2$	(12)	$S_2 \times S_3$	$\frac{120}{2\times 6} = 10$
$1+1+3$	(123)	$\mathbf{Z}_3 \times S_2$	$\frac{120}{3\times 2} = 20$
$1+2+2$	$(12)(34)$	$(\mathbf{Z}_2)^2 \rtimes \mathbf{Z}_2$	$\frac{120}{4\times 2} = 15$
$1+4$	(1234)	$\mathbf{Z}_4$	$\frac{120}{4} = 30$
$2+3$	$(12)(345)$	$S_2 \times \mathbf{Z}_3$	$\frac{120}{2\times 3} = 20$
5	(12345)	$\mathbf{Z}_5$	$\frac{120}{5} = 24$

The following is an immediate application of class equation.

Definition 6.3.3. Let p be a prime number. A finite group G is called a **p-group** if $|G|$ is a power of p.

Proposition 6.3.4. *For a nontrivial p-group G, $Z(G)$ is nontrivial.*

Proof. We use class formation for the p-group G:

$$\underset{\substack{\uparrow \\ p\text{-power}}}{\boxed{|G|}} \;=\; |Z(G)| \;+\; \sum_{i=1}^{t} \underset{\substack{\uparrow \\ \text{nontrivial } p\text{-power}}}{\boxed{\big[G : C_G(g_i)\big]}}$$

From this, we see that p divides $|Z(G)|$. But $\{e\} \in Z(G)$, so $Z(G)$ is nontrivial. □

6.4. **Automorphisms of a group.** Let G be a group. Recall that

$$\mathrm{Aut}(G) = \big\{\phi : G \xrightarrow{\simeq} G \text{ isomorphism}\big\}$$

Recall that the conjugation gives a homomorphism

$$\begin{tikzcd}
\mathrm{Ad} : G \arrow[r] & \mathrm{Aut}(G) \\
g \arrow[r, mapsto] & \big(\mathrm{Ad}_g : h \mapsto ghg^{-1}\big).
\end{tikzcd}$$

We have shown in Properties 6.1.6(2) that $\ker(\mathrm{Ad}) = Z(G)$.

Definition 6.4.1. The subgroup $\mathrm{Ad}(G) \subseteq \mathrm{Aut}(G)$, denoted by $\mathrm{Inn}(G)$, are called the group of **inner automorphisms**.

The following lemma will show that $\mathrm{Inn}(G)$ is a normal subgroup of $\mathrm{Aut}(G)$. The quotient

$$\mathrm{Out}(G) := \mathrm{Aut}(G)/\mathrm{Inn}(G)$$

is called the group of **outer automorphisms** of G.

Lemma 6.4.2. *For a group G, $\mathrm{Inn}(G) \lhd \mathrm{Aut}(G)$ is a normal subgroup.*

Proof. We need to show that, if $\sigma : G \xrightarrow{\simeq} G$ is an automorphism, then $\sigma \mathrm{Inn}(G)\sigma^{-1} = \mathrm{Inn}(G)$. (In fact, it suffices to prove "$\subseteq$", and "$\supseteq$" follows from the inclusion "$\subseteq$ for σ^{-1}".)

For this, take $\mathrm{Ad}_g \in \mathrm{Inn}(G)$ for $g \in G$. We claim that

$$\sigma \circ \mathrm{Ad}_g \circ \sigma^{-1} : G \to G$$

as an automorphism is equal to $\mathrm{Ad}_{\sigma(g)}$, so it belongs to $\mathrm{Inn}(G)$; and thus proving $\sigma\mathrm{Inn}(G)\sigma^{-1} \subseteq \mathrm{Inn}(G)$. Indeed, we have

$$\begin{aligned} \sigma \circ \mathrm{Ad}_g \circ \sigma^{-1}(h) &= \sigma\big(\mathrm{Ad}_g(\sigma^{-1}(h))\big) = \sigma\big(g\sigma^{-1}(h)g^{-1}\big) \\ &= \sigma(g)\sigma(\sigma^{-1}(h))\sigma(g)^{-1} = \sigma(g)h\sigma(g)^{-1} = \mathrm{Ad}_{\sigma(g)}(h). \end{aligned}$$

□

Example 6.4.3. Consider $G = \mathrm{GL}_n(\mathbb{Q})$, the conjugation action gives $\mathrm{Ad} : \mathrm{GL}_n(\mathbb{Q}) \to \mathrm{Aut}(G)$, then

$$\begin{aligned} \ker(\mathrm{Ad}) = Z(\mathrm{GL}_n(\mathbb{Q})) &= \big\{A \in \mathrm{GL}_n(\mathbb{Q}) \,\big|\, AB = BA, \text{ for all } B \in \mathrm{GL}_n(\mathbb{Q})\big\} \\ &= \big\{a \cdot I_n \,\big|\, a \in \mathbb{Q}^\times\big\} \cong \mathbb{Q}^\times. \end{aligned}$$

Thus we have $\mathrm{Inn}(G) \cong \mathrm{GL}_n(\mathbb{Q})/\mathbb{Q}^\times =: \mathrm{PGL}_n(\mathbb{Q})$ (the **projective general linear group**).

What about automorphisms that are not inner? The automorphism

$$\begin{aligned} \psi : \mathrm{GL}_n(\mathbb{Q}) &\longrightarrow \mathrm{GL}_n(\mathbb{Q}) \\ A &\longmapsto {}^tA^{-1} \end{aligned}$$

satisfies $\psi(AB) = \psi(A)\psi(B)$. This gives an action by automorphism

$$\mathrm{PGL}_n(\mathbb{Q}) \rtimes \{1, \psi\} \curvearrowright \mathrm{GL}_n(\mathbb{Q})$$

Remark 6.4.4. Write $\mathrm{SL}_n(\mathbb{Q}) := \{A \in \mathrm{GL}_n(\mathbb{Q}) \mid \det A = 1\}$. Then it is known that

$$\mathrm{Aut}(\mathrm{SL}_n(\mathbb{Q})) \cong \mathrm{Aut}(\mathrm{PGL}_n(\mathbb{Q})) \cong \begin{cases} \mathrm{PGL}_n(\mathbb{Q}) \rtimes \{1, \psi\} & \text{when } n \geq 3 \\ \mathrm{PGL}_2(\mathbb{Q}) & \text{when } n = 2. \end{cases}$$

Here the reason that we do not have ψ when $n = 2$ is that in this case $\psi : \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto {}^t\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)^{-1} = \left(\begin{smallmatrix} d & -c \\ -b & a \end{smallmatrix}\right)$ is the same as conjugation by $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. (One way to explain this is that: there are two ways to think of the projective space $\mathbb{P}^{n-1}$, a line in $\mathbb{Q}^n$ or a hyperplane in $\mathbb{Q}^n$; there is some kind of duality to take one version into the other; this is reflected by ψ. But when $n = 2$, the two types of view are the same; so ψ is an inner automorphism.)

What is $\mathrm{Aut}(\mathrm{GL}_n(\mathbb{Q}))$? The issue is the center $\mathbb{Q}^\times$. Abstractly $\mathbb{Q}^\times \cong \{\pm 1\} \times \prod\limits_{p \text{ prime}} p^{\mathbb{Z}}$; so it is a huge infinite product, and has very large automorphisms. Of course, such automorphisms are not very interesting, if we confine to automorphisms of G that are "given by polynomials maps", then (when $n \geq 3$)

$$\mathrm{Aut}(\mathrm{GL}_n(\mathbb{Q}))^{\mathrm{alg}} \cong \mathrm{PGL}_n(\mathbb{Q}) \rtimes \{1, \psi\}.$$

Example 6.4.5. For $G = S_n$, the conjugation action $\mathrm{Ad} : S_n \to \mathrm{Aut}(S_n)$ is injective. Here is an interesting fact:

If $n \neq 6$, $\mathrm{Ad} : S_n \to \mathrm{Aut}(S_n)$ is an isomorphism, i.e. all automorphisms of S_n are "inner".

When $n = 6$, there exists an automorphism $\psi : S_6 \xrightarrow{\simeq} S_6$ that is *not* inner, given by

$$\psi((12)) = (12)(34)(56), \quad \psi((23)) = (14)(25)(36), \quad \psi((34)) = (13)(24)(56),$$
$$\psi((45)) = (12)(36)(45), \quad \text{and} \quad \psi((56)) = (14)(23)(56).$$

In fact, one can prove that $\mathrm{Aut}(S_6) = S_6 \rtimes \{1, \psi\}$.

Extended Readings after Section 6

6.5. Characteristic subgroups.

Definition 6.5.1. Let G be a group. We say that a subgroup H is **characteristic**, denoted as $H \operatorname{char} G$, if for any automorphism σ of G, $\sigma(H) = H$.

Properties 6.5.2.

(1) If $H \leq G$ is the unique subgroup of that order, then H is characteristic.

For example, in $\mathbf{Z}_n$, for every $d|n$, $\langle d \rangle \subseteq \mathbf{Z}_n$ is characteristic.

(2) Characteristic subgroups are normal.

Indeed, if $H \operatorname{char} G$ and $g \in G$, $\mathrm{Ad}_g : G \to G$ is an automorphism. The definition of characteristic subgroups gives:

$$\mathrm{Ad}_g(H) = H \;\Rightarrow\; H \trianglelefteq G.$$

(3) If $K \operatorname{char} H$ and $H \trianglelefteq G$, then $K \trianglelefteq G$.

Moreover, if $K \operatorname{char} H$ and $H \operatorname{char} G$, then $K \operatorname{char} G$, i.e. characteristic subgroup is transitive.

(Proving the first statement: given any $g \in G$, as $H \trianglelefteq G$, we have $gHg^{-1} = H$. Thus

$$\begin{aligned} \mathrm{Ad}_g : H &\longrightarrow H \\ h &\longmapsto ghg^{-1} \end{aligned}$$

is an automorphism. The property of characteristic subgroups implies that $\mathrm{Ad}_g(K) = K$.) The proof of the second statement is left to the readers.

7. Sylow's theorems

In this lecture, we focus on one of the most important tools in the study of finite groups: Sylow's theorems. In some sense, the main motivation of Sylow's theorem is to find abstract ways to

- study groups in a way similar to, for $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, having $\mathbf{Z}_n = \mathbf{Z}_{p_1^{\alpha_1}} \times \cdots \times \mathbf{Z}_{p_r^{\alpha_r}}$, and hence reducing the study of $\mathbf{Z}_n$ to each of prime p; (namely to separating the structure to the part for each prime p) and
- consider the group G acting on some natural set X (as a way to *re-present* the group).

7.1. Statement of Sylow's theorems.

Definition 7.1.1. Fix a prime number p.

(1) A **p-group** is a finite group whose order is a power of p.
(2) If G is a finite group of order $|G| = p^r m$ with $r, m \in \mathbb{N}$ and $p \nmid m$, a subgroup H of G of order exactly p^r is called a **Sylow p-subgroup** or **p-Sylow subgroup**. Write

$$\mathrm{Syl}_p(G) := \{\text{Sylow } p\text{-subgroups of } G\} \quad \text{and} \quad n_p := \big|\mathrm{Syl}_p(G)\big|.$$

Theorem 7.1.2 (Sylow's theorem)**.** *Let G be a finite group with $|G| = p^r m$ with $r, m \in \mathbb{N}$ and $p \nmid m$.*

- (First Sylow Theorem) *Sylow p-subgroups exist.*
- (Second Sylow Theorem) *If P is a Sylow p-subgroup of G, and $Q \leq G$ is a subgroup of p-power order, then there exists $g \in G$ such that $Q \leq gPg^{-1}$ (note gPg^{-1} is also a Sylow p-subgroup).*
  In other words, we have
  – *all Sylow p-subgroups are conjugate; and*
  – *all subgroups of p-power order is contained in a Sylow p-subgroup.*
- (Third Sylow Theorem) *The number $n_p = \big|\mathrm{Syl}_p(G)\big|$ satisfies*
  (1) *$n_p \equiv 1 \bmod p$, and*
  (2) *$n_p | m$.*

7.2. Proof of Sylow's theorems and their corollaries.

7.2.1. *Proof of First Sylow Theorem.* (There are a few proofs of this theorem. The one we include here can be found in Dummit–Foote's book, which in my opinion requires some "high-level thinking" but less trickier; we hope to give the readers a sense of how a "structured proof" might look like as opposed to a proof only involves tricks. In the extended material, we include another proof.)

We use an induction on $|G|$. When $|G| = 1$, there is nothing to prove.

Suppose that the First Sylow Theorem is proved for finite groups of order $< n$. Let G be a finite group of order $n = p^r m$ with $r, m \in \mathbb{N}$ and $p \nmid m$.

Case 1: $p \nmid |G|$, i.e. $r = 0$. Then $\{1\} \subseteq G$ is the Sylow p-subgroup of G.

Case 2: If p divides $|Z(G)|$, then $Z(G)$ is a finitely generated abelian group; so

$$Z(G) = \underbrace{\mathbf{Z}_{p^{r_1}} \times \cdots \times \mathbf{Z}_{p^{r_s}}}_{p\text{-part}} \times \cdots$$

We write $Z(G)_p$ for the p-part of $Z(G)$; then $|Z(G)_p| = p^{r'}$ for some $r' \geq 1$.

Now, we consider the quotient homomorphism $G \xrightarrow{\pi}\!\!\!\!\rightarrow G/Z(G)_p := \overline{G}$, where the quotient $\overline{G}$ has order $n/p^{r'} = p^{r-r'}m < n$. By inductive hypothesis, $\overline{G}$ contains a Sylow p-subgroup $\overline{H}$ of order $p^{r-r'}$. Then $\pi^{-1}(\overline{H})$ is a subgroup of G of order

$$|\overline{H}| \cdot |\ker \pi| = p^{r-r'} \cdot p^{r'} = p^r.$$

So H is a Sylow p-subgroup of G.

Case 3: If p does not divide $|Z(G)|$ but p divides $|G|$ (thus $r \geq 1$).

We use the class equation for the conjugation of G on itself proved in the previous lecture (Theorem 6.3.1), to deduce that

$$\underset{\underset{\text{div. by } p}{\uparrow}}{\boxed{|G|}} \quad = \quad \underset{\underset{\underline{\text{not}}\text{ div. by } p}{\uparrow}}{\boxed{|Z(G)|}} \quad + \quad \underset{\underset{\text{sum over representatives of conjugacy classes}}{\uparrow}}{\sum_{i=1}^{t} \big[G : C_G(g_i)\big]}$$

It follows that there exists one i such that $[G : C_G(g_i)]$ is *not* divisible by p. Thus $C_G(g_i)$ has order $p^r m'$ for some $m'|m$ and $m' \neq m$.

By inductive hypothesis applied to $C_G(g_i)$, there exists a subgroup H of $C_G(g_i)$ of order p^r. This H is also a Sylow p-subgroup of G.

This completes the inductive proof of First Sylow Theorem. □

7.2.2. *Proof of Second Sylow Theorem.* Now let $P \leq G$ be a Sylow p-subgroup, and $Q \leq G$ a subgroup of p-power order.

When $|Q| = 1$, i.e. $Q = \{1\}$, clearly, $Q \leq P$, we are done.

Now we assume that $|Q| = p^{r'}$ with $r' \geq 1$. Consider the left translation action of Q on G/P:

$$Q \circlearrowright G/P = \big\{gP \,\big|\, g \in G\big\}.$$

More precisely, $q \cdot gP := qgP$, for $q \in Q$ and $gP \in G/P$.

Then we must have

$$\underset{\underset{\text{not divisible by } p}{\uparrow}}{\boxed{|G/P|}} \quad = \sum_{i=1}^{t} \big|\text{Orbits}_i\big| = \sum_{i=1}^{t} \big|Q/\text{Stab}_i\big|$$

Thus, there exists one orbit, say the orbit of gP such that the number of elements in the orbit is not divisible by p. Let

$$Q' := \big\{q \in Q \,\big|\, qgP = gP\big\} = \text{Stab}_Q(gP) \leq Q$$

be the stabilizer group. It follows that $|Q/Q'|$ is the same as the size of the orbit, and is thus not divisible by p.

Yet $|Q| = p^{r'}$ is a power of p, we must have $Q = Q'$. In particular, this says that for any $q \in Q$, we have

$$qgP = gP \quad \Rightarrow \quad qg \in gP \quad \Rightarrow \quad q \in gPg^{-1}.$$

So we deduce that $Q \leq gPg^{-1}$.

Corollary 7.2.3. *All Sylow subgroups are conjugate.*

Proof. Let P and Q be two Sylow p-subgroups. The Second Sylow Theorem implies that there exists $g \in G$ such that $Q \leq gPg^{-1}$. But $|Q| = |gPg^{-1}|$. So $Q = gPg^{-1}$. □

Corollary 7.2.4. *There is only one Sylow p-subgroup if and only if one Sylow p-subgroup $P \leq G$ is normal.*

Proof. "$\Rightarrow$": for any $g \in G$, gPg^{-1} is also a Sylow p-subgroup. So the uniqueness implies that $P = gPg^{-1}$. Thus P is normal.

"$\Leftarrow$": Since all Sylow p-subgroups are conjugate by the above Corollary and all conjugates gPg^{-1} are the same as P, there is only one Sylow p-subgroup, namely P. □

Remark 7.2.5. Usually, when quoting Corollaries 7.2.3 and 7.2.4, we will just say Second Sylow Theorem.

Corollary 7.2.6. *If P is a Sylow p-subgroup, then $N_G(N_G(P)) = N_G(P)$, and $N_G(P)$ contains a unique Sylow p-subgroup, which is P.*

Proof. Note that $P \trianglelefteq N_G(P)$ tautologically holds; so P is a normal Sylow p-subgroup of $N_G(P)$. By the above Corollary, P is the unique Sylow p-subgroup of $N_G(P)$. (This proves the second statement.)

It is clear that $N_G(P) \subseteq N_G(N_G(P))$. Conversely, for $n \in N_G(N_G(P))$, we have

$$\begin{array}{ccc} nN_G(P)n^{-1} & = & N_G(P) \\ \bigcup & & \bigcup \\ nPn^{-1} & & P \end{array}$$

But we have just shown that $N_G(P)$ has a unique Sylow p-subgroup. So $nPn^{-1} = P$, i.e. $n \in N_G(P)$. □

7.2.7. *Proof of Third Sylow Theorem.* Recall that $|G| = n = p^r m$ with $r \in \mathbb{Z}_{\geq 0}$ and $p \nmid m$. Put $n_p := \big|\mathrm{Syl}_p(G)\big|$.

(1) Consider the conjugation of G on $\mathrm{Syl}_p(G)$: $g \star P := gPg^{-1}$ for $g \in G$ and P a Sylow p-subgroup of G.

By Second Sylow Theorem, all Sylow p-subgroups are conjugate and thus the conjugation G-action on $\mathrm{Syl}_p(G)$ is transitive. From this, we deduce

$$n_p = \big|\mathrm{Syl}_p(G)\big| = \big|G/N_G(P)\big| = \frac{|G|}{|N_G(P)|} = \frac{p^r \cdot m}{p^r \cdot [N_G(P) : P]}.$$

It is clear from this that $n_p | m$.

(2) Consider P acting on $\mathrm{Syl}_p(G)$ by conjugation. Then we have

$$n_p = \big|\mathrm{Syl}_p(G)\big| = \sum_{\text{orbits } \mathrm{Ad}_P(P_i)} \big|P/\mathrm{Stab}_P(P_i)\big|. \qquad (7.2.7.1)$$

- If $\mathrm{Stab}_P(P_i) \neq P$, then $\mathrm{Stab}_P(P_i)$ is a subgroup of P; Lagrange theorem implies that $\big|P/\mathrm{Stab}_P(P_i)\big|$ is a nontrivial p-power and in particular divisible by p.
- If $\mathrm{Stab}_P(P_i) = P$, then $P \subseteq N_G(P_i)$. But Corollary 7.2.6 implies that $N_G(P_i)$ contains a unique Sylow p-subgroup, i.e. $P = P_i$. So there is a unique such orbit.

It follows from the above discussion that $n_p \equiv 1 \pmod{p}$.

7.3. **Applications of Sylow's Theorem.**

7.3.1. *Classification of groups of order pq.* Assume that G is a group of order pq, where $p < q$ are prime numbers.

Let Q denote a Sylow q-subgroup of G. By Sylow's Third Theorem, we have

$$n_q \,\big|\, p, \quad \text{and} \quad n_q \equiv 1 \pmod q \qquad \Rightarrow \qquad n_q = 1.$$

So Q is a normal subgroup (and it is unique).

Let P be a Sylow p-subgroup. The Sylow's Third Theorem says that

$$n_p \,\big|\, q \quad \text{and} \quad n_p \equiv 1 \pmod p.$$

Thus $n_p = 1$ or q. We will separate two cases, but in either case, we must have $P \simeq \mathbf{Z}_p$ and $Q \simeq \mathbf{Z}_q$.

Case 1 : $n_p = 1$. Then P is a normal subgroup. By Theorem 5.1.1, we have $G = P \times Q$.

Case 2 : $n_p = q \equiv 1 \pmod p$. In this case, $G = QP$ (because $P \cap Q = \{1\}$). As Q is a normal subgroup and $P \cap Q = \{1\}$, Proposition 5.4.5 implies that $G = Q \rtimes P$. This corresponds to a homomorphism

(7.3.1.1)
$$\begin{tikzcd} \mathbf{Z}_p \simeq P \arrow[r, "\eta"] & \mathrm{Aut}(Q) \cong \mathrm{Aut}(\mathbf{Z}_q) \arrow[r, phantom, "\cong"] & \mathbf{Z}_q^\times \\ y \arrow[r, mapsto] & (x \mapsto yxy^{-1} = y^a) \arrow[r, mapsto] & a. \end{tikzcd}$$

Then $G \simeq Q \rtimes_\eta P$.

Next, we claim that, up to isomorphism, there is only one such semidirect product.

We use the fact that $\mathbf{Z}_q^\times$ is a cyclic group of order $\varphi(q) = q - 1$, i.e. $\mathbf{Z}_q^\times \simeq \mathbf{Z}_{q-1}$. Then there is a *unique* subgroup or order p of $\mathbf{Z}_{q-1}$. Written as a subgroup of $\mathbf{Z}_q^\times$, it takes the form of

$$\{1,\, a,\, \ldots,\, a^{p-1}\}.$$

Fix a generator $\sigma \in P$ of P. Then all possible such homomorphism (7.3.1.1) is of the form (for $i = 1, \ldots, p-1$):

$$\begin{tikzcd} \eta_i : P = \mathbf{Z}_p \arrow[r] & \mathbf{Z}_q^\times \\ \sigma \arrow[r, mapsto] & a^i. \end{tikzcd}$$

We then define an isomorphism:

$$\begin{tikzcd} Q \rtimes_{\eta_1} \mathbf{Z}_p \arrow[r, "\simeq"] & Q \rtimes_{\eta_i} \mathbf{Z}_p \\ (y, c) \arrow[r, mapsto] & (y, ic). \end{tikzcd}$$

7.3.2. *A group of order* 132 *cannot be a simple group.* (This is a typical application of Sylow's theorem.)

We factor $132 = 11 \times 3 \times 4$. Suppose that G is simple, then G would not have any nontrivial normal subgroups. In particular, n_2, n_3, and n_{11} are not 1. In any case, the Sylow p-subgroup P_p for $p = 3, 11$ has order p, and hence is isomorphic to $\mathbf{Z}_p$.

- Consider n_{11}. $n_{11} \equiv 1 \bmod 11$ and $n_{11} | 12$. So $n_{11} = 12$. As each pair of Sylow 11-subgroups have only trivial intersection, there are at least $12 \cdot 10 = 120$ elements of order 11.

- Consider n_3. $n_3 \equiv 1 \bmod 3$ and $n_3|44$. So $n_3 = 4, 22$. Using the same argument as above, we see that there are at least $2 \times 4 = 8$ elements of order exactly 3.
- Consider $n_2 \geq 2$. The intersection of two Sylow 2-subgroup has at most 2 elements. So there are at least $4 + 4 - 2 - 1 = 7$ elements whose order is 2-power.

In total, there are at least $120 + 8 + 7 = 135$ elements in G. This is a contradiction.

7.3.3. *Sylow groups for normal subgroups and quotient groups.* Consider the case of a finite group G and a normal subgroup N. Let $\pi : G \twoheadrightarrow G/N$ denote the projection (with kernel $\ker \pi = N$).

We claim that, for a Sylow p-subgroup H of G,

(1) the image $\pi(H)$ is a Sylow p-subgroup of G/N; and
(2) the intersection $N \cap H$ is a Sylow p-subgroup of N.

We make the following observation: restricting π to the subgroup H, we obtain a map

$$\pi|_H : H \twoheadrightarrow \pi(H).$$

The kernel is $\ker(\pi|_H) = N \cap H$. By First Isomorphism Theorem, we have

$$\pi(H) \cong H\big/(N \cap H).$$

Recall our convention: $|G| = p^r m$ with $r \in \mathbb{Z}_{\geq 0}$ and $p \nmid m$. Put $|N| = p^{r'}m'$ with $r' \in \mathbb{Z}_{\geq 0}$, $r' \leq r$, and $m'|m$. Then $|G/N| = p^{r-r'}\frac{m}{m'}$.

Yet, $N \cap H$ is a subgroup of N and $\pi(H)$ is a subgroup of G/N. Lagrange theorem gives:

$$\begin{tikzcd}[column sep=small] & {|N \cap H| \text{ divides } |N| = p^{r'}m'} \arrow[dl, squiggly] \\ {p^r = |H| = |N \cap H| \cdot |\pi(H)|} & \\ & {|\pi(H)| \text{ divides } |G/N| = p^{r-r'}\frac{m}{m'}.} \arrow[ul, squiggly] \end{tikzcd}$$

From this, we see that we are forced to have $|N \cap H| = p^{r'}$ and $|\pi(H)| = p^{r-r'}$. This shows that $\pi(H)$ is a Sylow p-subgroup of G/N and $N \cap H$ is a Sylow p-subgroup of N.

7.3.4. *Group of order* 105. Let G be a group of order 105 containing a normal Sylow 3-subgroup P_3. We will show that $G \simeq \mathbf{Z}_{105}$.

We start by considering Sylow 5-subgroups and Sylow 7-subgroups.

Sylow's Third Theorem says that

$$n_5 \,\big|\, 3 \cdot 7 = 21, \quad n_5 \equiv 1 \pmod 5 \qquad \Longrightarrow n_5 = 1,\, 21.$$
$$n_7 \,\big|\, 3 \cdot 5 = 15, \quad n_7 \equiv 1 \pmod 7 \qquad \Longrightarrow n_7 = 1,\, 15.$$

We need some additional input to proceed (in particular, using the normality of P_3.)

Since P_3 is a normal subgroup of G, the conjugation action of G on P_3 defines a homomorphism

$$\begin{aligned} \phi : G &\longrightarrow \mathrm{Aut}(P_3) \simeq \mathbf{Z}_3^\times \simeq \mathbf{Z}_2 \\ g &\longmapsto (\mathrm{Ad}_g : x \mapsto gxg^{-1}) \end{aligned}$$

In particular, $\phi(G) \subseteq \mathbf{Z}_2$ is a subgroup (so $\phi(G) = \{1\}$ or $\phi(G) = \mathbf{Z}_2$. Yet $\phi(G)$ is a quotient of G, which has odd order; so $|\phi(G)|$ is odd, and thus $\phi(G) = 1$.

In other words, for any $g \in G$, $x = gxg^{-1}$ for any $x \in P_3$; this means that $P_3 \subseteq Z(G)$.

Now, we need to go back to the proof of Sylow Third Theorem (1), letting G acting on $\{$Sylow p-subgroups$\}$ (for $p = 5, 7$). Now P_3 acts trivially under conjugation; so n_5 is a factor of $|G/(P_3P_5)| = 7$. So $n_5 = 1$. Similar argument shows that $n_7 = 1$. So both P_5 and P_7 are normal subgroups.

Now, all P_3, P_5, and P_7 are normal; $G = P_3 \times P_5 \times P_7 \simeq \mathbf{Z}_{105}$.

Extended reading material

7.4. **An alternative proof of first Sylow's theorem.** We need the following.

Notation 7.4.1. For a positive integer n and a prime p, we write $v_p(n)$ for the maximal nonnegative integer m such that p^m divides n.

If we write $n = a_0 + a_1p + a_2p^2 + \cdots$, put $\mathrm{Dig}_p(n) := a_0 + a_1 + a_2 + \cdots$ for the sum of digits when writing n as a p-based number.

A general fact is that

$$v_p(n!) = \Big\lfloor \frac{n}{p} \Big\rfloor + \Big\lfloor \frac{n}{p^2} \Big\rfloor + \Big\lfloor \frac{n}{p^3} \Big\rfloor + \cdots = \frac{n - \mathrm{Dig}_p(n)}{p-1}.$$

The last equality can be checked as follows: it is enough to prove this for $n = a_ip^i$, for which it reads $a_i(p^{i-1} + p^{i-2} + \cdots + 1) = \dfrac{a_ip^i - a_i}{p-1}$.

Lemma 7.4.2. *Let p be a prime number and let $n = p^rm$ with $p \nmid m$. Then for any $0 \leq k \leq r$, we have*

$$v_p\Big(\binom{n}{p^k}\Big) = r - k.$$

Proof. We compute this as follows:

$$\begin{aligned} v_p\Big(\binom{n}{p^k}\Big) &= \frac{\big(n - \mathrm{Dig}_p(n)\big) - \big(p^k - \mathrm{Dig}_p(p^k)\big) - \big(n - p^k - \mathrm{Dig}_p(n-p^k)\big)}{p-1} \\ &= \frac{\mathrm{Dig}_p(n-p^k) - +\mathrm{Dig}_p(p^k) - \mathrm{Dig}(n)}{p-1} = r - k \end{aligned}$$

where the last equality is because it computes the number of times that we take over a p to the next digit when computing $(n - p^k) + p^k$. □

Theorem 7.4.3. *Let G be a group of order $n = p^rm$ with $p \nmid m$ and $r \in \mathbb{Z}_{\geq 0}$. Then G contains a Sylow p-subgroup.*

Proof. Let Ω denote the set of subsets of G of p^k elements, i.e. an element A of Ω is a subset $\{a_1, a_2, \dots, a_{p^k}\}$. We define a G-action on Ω by, for $g \in G$,

$$g \star A := \{ga_1, ga_2, \dots, ga_{p^k}\}.$$

We may write Ω as the disjoint union of orbits, so

$$|\Omega| = \sum_{\text{orbits } \mathcal{O}} |\mathcal{O}|.$$

Using Lemma 7.4.2, we see that

$$p^{r-k+1} \nmid |\Omega| = \binom{n}{p^k}.$$

So there exists at least one orbit $\mathcal{O} = G \cdot A$ such that $p^{r-k+1} \nmid |\mathcal{O}|$.

We claim that $\text{Stab}_G(A)$ is a Sylow p-subgroup. Indeed, we note that

$$|\mathcal{O}| = \frac{|G|}{|\text{Stab}_G(A)|} = \frac{p^r m}{|\text{Stab}_G(A)|}.$$

But $p^{r-k+1} \nmid |\mathcal{O}|$; so $p^k \mid |\text{Stab}_G(A)|$.

On the other hand, pick any element $a \in A$,

$$\{ga \mid g \in \text{Stab}_G(A)\} \subseteq A$$

and the elements ga are pairwise distinct. So $|\text{Stab}_G(A)| \leq |A| = p^k$. Combining this with above shows that $|\text{Stab}_G(A)| = p^k$. So $\text{Stab}_G(A)$ is a Sylow p-subgroup. □

8. Commutator subgroups, nilpotent groups, and p-groups

8.1. **Commutator subgroups.**

Definition 8.1.1. For $x, y \in G$, define $[x, y] := x^{-1}y^{-1}xy$, the **commutator** of x and y.

Let $G^{\mathrm{der}} = G' := \big\langle [x, y] \,\big|\, x, y \in G \big\rangle$ be the subgroup of G generated by all commutators; this is called the **commutator subgroup** or the **derived subgroup** of G.

Note: *It is NOT true that every element of G' is a commutator itself.*

Properties 8.1.2. The commutator of $x, y \in G$ enjoy the following properties:

(1) $xy = yx$ if and only if $[x, y] = 1$;
(2) for $g \in G$, we have $g[x, y]g^{-1} = [gxg^{-1}, gyg^{-1}]$;
(3) G' is a normal subgroup of G and G/G' is abelian.

Proof. (1) is clear. For (2), we compute directly

$$g[x, y]g^{-1} = gx^{-1}y^{-1}xyg^{-1} = gx^{-1}g^{-1} \cdot gy^{-1}g^{-1} \cdot gxg^{-1} \cdot gyg^{-1} = [gxg^{-1}, gyg^{-1}].$$

(3) For $g \in G$, $gG'g^{-1}$ is generated by elements of the form $g[x, y]g^{-1}$, which are the same as $[gxg^{-1}, gyg^{-1}]$ by (2). So $gG'g^{-1} = G'$.

Moreover, for $x, y \in G$, we have

$$xG' \cdot yG' = yG' \cdot xG' \;\Leftrightarrow\; x^{-1}y^{-1}xyG' = G'.$$

So xG' and yG' commute with each other in G/G'; it is abelian. □

Proposition 8.1.3. *If A is an abelian group and $\phi : G \to A$ is a homomorphism, then $G' \subseteq \ker \phi$. Moreover, ϕ factors as the composition of*

(8.1.3.1)
$$\begin{tikzcd} G \arrow[r, two heads, "\pi"] & G/G' \arrow[r, "\bar{\phi}"] & A \\ g \arrow[r, mapsto] & gG' \arrow[r, mapsto] & \phi(g). \end{tikzcd}$$

In particular, we have a bijection for every abelian group A:

$$\begin{tikzcd} \mathrm{Hom}_{\mathrm{gp}}(G, A) \arrow[r, "\cong"] & \mathrm{Hom}_{\mathrm{gp}}(G/G', A) \\ \phi \arrow[r, mapsto] & \big(\bar{\phi} : gG' \mapsto \phi(g)\big). \end{tikzcd}$$

In other words, if we want to Hom a group out to an abelian group, it is enough to Hom out from G/G'.

Proof. Note that for any $x, y \in G$, $\phi([x, y]) = \phi(x^{-1}y^{-1}xy) = \phi(x)^{-1}\phi(y)^{-1}\phi(x)\phi(y) = 1$ because A is abelian. So $G' \subseteq \ker \phi$. It follows that G must factor through G/G' as shown in (8.1.3.1). In particular this gives the map $\mathrm{Hom}_{\mathrm{gp}}(G, A) \to \mathrm{Hom}_{\mathrm{gp}}(G/G', A)$. The converse map is easier and it follows from sending a homomorphism $\psi : G/G' \to A$ to the composition

$$\begin{tikzcd} G \arrow[r, two heads, "\pi"] & G/G' \arrow[r, "\psi"] & A. \end{tikzcd}$$

□

Example 8.1.4. For $G = D_{2n} = \big\langle r, s \,\big|\, r^n = s^2 = 1,\, srs = r^{-1} \big\rangle$, compute all homomorphisms $\mathrm{Hom}_{\mathrm{gp}}(G, \mathbb{C}^\times)$.

First note that G' contains $srs^{-1}r^{-1} = r^{-2}$. We separate two cases.

- If n is odd, then $\langle r\rangle = \langle r^{-2}\rangle \subseteq G'$. We claim that $G' = \langle r\rangle$. Instead of checking every pair of elements, we use the following argument: construct a map
$$\begin{aligned} \psi : G &\longrightarrow \{\pm 1\} \\ r^i &\longmapsto 1 \\ sr^i &\longmapsto -1. \end{aligned}$$
One checks that $\psi(r)^n = \psi(s)^2 = 1$ and $\psi(s)\psi(r)\psi(s) = \psi(r)^{-1}$; so ψ is a homomorphism.

  By Proposition 8.1.3 above, we have $G' \subseteq \ker\psi = \langle r\rangle$. So $G' = \langle r\rangle$ and $G/G' \cong \{\pm 1\}$.
- If n is even, then $\langle r^{-2}\rangle = \langle r^2\rangle \subseteq G'$. Similar to above, we define
$$\begin{aligned} \psi : G &\longrightarrow \{\pm 1\} \times \pm 1\} \\ r^i &\longmapsto \big((-1)^i,\, 1\big) \\ s &\longmapsto \big(1,\, -1\big). \end{aligned}$$
Once again, we check the relations $\psi(r)^n = ((-1)^n, 1) = (1,1)$ as n is even, $\psi(s)^2 = (1,1)$, and $\psi(s)\psi(r)\psi(s) = (-1,1) = \psi(r)^{-1}$. So ψ is a well-defined homomorphism. Now Proposition 8.1.3 implies that $G' \subseteq \ker\psi = \langle r^2\rangle$. This implies that $G' = \langle r^2\rangle$ and $G/G' \cong \{\pm 1\} \times \{\pm 1\}$.

Now, we apply Proposition 8.1.3, to get:

- when n is odd,
$$\mathrm{Hom}_{\mathrm{gp}}(D_{2n}, \mathbb{C}^\times) \cong \mathrm{Hom}_{\mathrm{gp}}(\{\pm 1\}, \mathbb{C}^\times) \cong \big\{\mathrm{tr},\, \psi\big\}.$$
There are two such homomorphisms: the trivial one, and the ψ above, given by sending r to 1 and s to -1.
- when n is even,
$$\begin{tikzcd} \mathrm{Hom}_{\mathrm{gp}}(G, \mathbb{C}^\times) \arrow[r, leftrightarrow, "\cong"] & \mathrm{Hom}_{\mathrm{gp}}\big(\{\pm 1\} \times \{\pm 1\}, \mathbb{C}^\times\big) \\ \begin{matrix} \psi : G \to \mathbb{C}^\times \\ \psi(r) = \lambda,\ \psi(s) = \mu \end{matrix} & \begin{matrix} \bar{\psi}(-1,1) = \lambda \in \{\pm 1\} \subset \mathbb{C}^\times \\ \bar{\psi}(1,-1) = \mu \in \{\pm 1\} \subset \mathbb{C}^\times \end{matrix} \arrow[l, squiggly] \end{tikzcd}$$

8.2. **Solvable groups.** We recall from Definition 3.4.4 that a group G is called solvable if there exists a chain of subgroups $1 = G_0 \leq G_1 \leq \cdots \leq G_r = G$ such that $G_{i-1} \trianglelefteq G_i$ and G_i/G_{i-1} is abelian, for every $i = 1, \dots, r$. (When G is finite, this is equivalent to existing such a chain of subgroups such that each G_i/G_{i-1} is isomorphic to $\mathbf{Z}_{p_i}$ for some prime number p_i.)

In particular, all abelian groups are solvable.

A good way to test solvable groups is through the following.

Definition 8.2.1. For any group G, define the following sequence of subgroups inductively.
$$G^{(0)} = G, \quad G^{(1)} = [G, G], \quad G^{(i+1)} = [G^{(i)}, G^{(i)}] \text{ for any } i.$$
This is called the **derived** or **commutator series** of G.

An interesting typical example of commutator series is the following.

Example 8.2.2. Consider

$$G = \begin{pmatrix} \mathbb{C}^\times & \mathbb{C} & \mathbb{C} \\ 0 & \mathbb{C}^\times & \mathbb{C} \\ 0 & 0 & \mathbb{C}^\times \end{pmatrix} \supseteq G^{(1)} = \begin{pmatrix} 1 & \mathbb{C} & \mathbb{C} \\ 0 & 1 & \mathbb{C} \\ 0 & 0 & 1 \end{pmatrix} \supseteq G^{(2)} = \begin{pmatrix} 1 & 0 & \mathbb{C} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \supseteq G^{(3)} = \{I_3\}.$$

Proposition 8.2.3. *A group is solvable if and only if $G^{(n)} = \{1\}$ for some finite $n \in \mathbb{N}$.*

Proof. "$\Leftarrow$" Note that each $G^{(i)}$ is a normal subgroup of $G^{(i-1)}$ and the quotient $G^{(i-1)}/G^{(i)}$ is abelian. So

$$\{1\} = G^{(n)} \trianglelefteq G^{(n-1)} \trianglelefteq \cdots \trianglelefteq G^{(1)} \trianglelefteq G$$

is the required chain of subgroups with abelian subquotients.

"$\Rightarrow$" As G is solvable, there exists a chain of subgroups

$$\{1\} = H_0 \leq H_1 \leq H_2 \leq \cdots \leq H_r = G$$

such that $H_{i-1} \trianglelefteq H_i$ and H_i/H_{i-1} is abelian.

It follows that $[H_i, H_i] \subseteq H_{i-1}$. This implies that

$$G^{(1)} = [G, G] \subseteq H_{r-1},$$
$$G^{(2)} = [G^{(1)}, G^{(1)}] \subseteq [H_{r-1}, H_{r-1}] \subseteq H_{r-2},$$
$$\cdots \quad \cdots$$
$$G^{(i)} \subseteq H_{r+1-i}, \quad \cdots$$

Eventually, we prove $G^{(r+1)} \subseteq H_0 = \{1\}$. □

Remark 8.2.4.
(1) The derived series is the "fastest-decreasing" series so that the subquotients are abelian.
(2) The smallest $n \in \mathbb{Z}_{\geq 0}$ for which $G^{(n)} = \{1\}$ is called the **solvable length** of G.

Lemma 8.2.5. *All $G^{(i)}$ are normal subgroups of G. In fact, they are characteristic subgroups of G.*

Proof. Recall that $G^{(1)} = [G, G] = \langle x^{-1}y^{-1}xy \,|\, x, y \in G \rangle$. If $\phi : G \to G$ is an automorphism, we have

$$\begin{aligned} \phi(G^{(1)}) &= \langle \phi(x^{-1}y^{-1}xy) \,|\, x, y \in G \rangle \\ &= \langle \phi(x)^{-1}\phi(y)^{-1}\phi(x)\phi(y) \,|\, x, y \in G \rangle = G^{(1)}. \end{aligned}$$

Inductively, we prove that

$$\phi(G^{(i)}) = \phi([G^{(i-1)}, G^{(i-1)}]) = \big[\phi(G^{(i-1)}), \phi(G^{(i-1)})\big] = \big[G^{(i-1)}, G^{(i-1)}\big] = G^{(i)}.$$

Thus $G^{(i)}$ is characteristic and thus normal. □

Properties 8.2.6.
(1) If $H \leq G$, we must have $H^{(i)} \leq G^{(i)}$. So if G is solvable, then H is solvable.
(2) Let $\phi : G \to K$ be a surjective homomorphism. Then $\phi(G^{(i)}) = K^{(i)}$. So if G is solvable, then K is solvable.
(3) If $N \trianglelefteq G$ is a normal subgroup and both N and G/N are solvable, then G is also solvable.

8.3. **Nilpotent subgroups.** We introduce a notion of nilpotent group, which sits in the following list of groups with properties.

$$\{\text{cyclic groups}\} \subseteq \{\text{abelian groups}\} \subseteq \{\text{nilpotent groups}\} \subseteq \{\text{solvable groups}\} \subseteq \{\text{all groups}\}$$

Definition 8.3.1. For a group G, define the following subgroups

$$G^0 = G,\ G^1 = [G,G],\ G^{i+1} = [G, G^i] \text{ for } i,$$

Then we have a chain of subgroups $G^0 \geq G^1 \geq G^2 \geq \cdots$. This is called the **lower central series** of G.

Similar to Lemma 8.2.5, we may prove that each G^i is a normal subgroup of G. It is also clear that $G^i \geq G^{(i)}$.

The group G is called **nilpotent** if $G^c = \{1\}$ for some $c \in \mathbb{N}$. The smallest such c is called the **nilpotence class** of G.

Remark 8.3.2. The construction of lower central series commutes with passing to quotients, i.e. if $\pi : G \to H = G/N$ is the homomorphism given by taking quotient by a normal subgroup N, then $\pi(G^i) = H^i$.

Corollary 8.3.3. *If G is nilpotent, then G is solvable.*

Proof. If $G^c = \{1\}$ for some $c \in \mathbb{N}$, then $G^{(c)} \leq G^c = \{1\}$. So $G^{(c)} = 1$. □

Example 8.3.4. Going back to Example 8.2.2 For the groups

$$G = \begin{pmatrix} \mathbb{C}^\times & \mathbb{C} & \mathbb{C} \\ 0 & \mathbb{C}^\times & \mathbb{C} \\ 0 & 0 & \mathbb{C}^\times \end{pmatrix} \supseteq G^{(1)} = \begin{pmatrix} 1 & \mathbb{C} & \mathbb{C} \\ 0 & 1 & \mathbb{C} \\ 0 & 0 & 1 \end{pmatrix} \supseteq G^{(2)} = \begin{pmatrix} 1 & 0 & \mathbb{C} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \supseteq G^{(3)} = \{I_3\},$$

G is NOT nilpotent, because a direct computation shows that $G^{(1)} = G^1$ but $G^2 = [G, G^1] = G^1$ and all $G^c = G^1$ for $c \geq 1$.

On the other hand, $G^{(1)}$ and $G^{(2)}$ are nilpotent groups.

We also introduce a "dual picture".

Definition 8.3.5. For any group G, define the following subgroups

$$Z_0(G) = 1, \quad Z_1(G) = Z(G).$$

Consider the following quotient

$$\begin{tikzcd} G \arrow[r, two heads, "\pi"] & G/Z(G) \\ Z_2(G) \arrow[u, phantom, "\bigcup" description] \arrow[r, two heads] & Z\big(G/Z(G)\big) \arrow[u, phantom, "\bigcup" description] \end{tikzcd}$$

where $Z_2(G)$ is defined to be $\pi^{-1}(Z(G/Z(G)))$. Since $Z(G/Z(G))$ is a normal subgroup of $G/Z(G)$, $Z_2(G)$ is a normal subgroup of G.

Now, inductively let $Z_{i+1}(G)$ be the subgroup of G containing $Z_i(G)$ such that $Z_{i+1}(G)/Z_i(G) \cong Z\big(G/Z_i(G)\big)$.

$$\begin{tikzcd} G \arrow[r, two heads, "\pi"] & G/Z_i(G) \\ Z_{i+1}(G) \arrow[u, phantom, "\bigcup" description] \arrow[r, two heads] & Z\big(G/Z_i(G)\big) \arrow[u, phantom, "\bigcup" description] \end{tikzcd}$$

(In particular, here, we always have inductively $Z_i(G) \trianglelefteq G$ is a normal subgroup. Suppose that $Z_i(G)$ is a normal subgroup. Yet $Z\big(G/Z_i(G)\big)$ is a normal subgroup of $G/Z_i(G)$; so its preimage $Z_{i+1}(G) := \pi^{-1}\big(Z\big(G/Z_i(G)\big)\big)$ is a normal subgroup of G.

The sequence $\{1\} = Z_0(G) \leq Z_1(G) \leq \cdots$ is called the **upper central series** of G.

Remark 8.3.6. We remark on the convention in choosing superscript versus subscript: typically, the convention is

- indexing by subscript for increasing filtration,
- indexing by superscript for decreasing filtration.

Theorem 8.3.7. *A group G is nilpotent if and only if $Z_c(G) = G$ for some $c \in \mathbb{N}$.*

More precisely, for some $c \in \mathbb{N}$, $G^c = \{1\}$ if and only if $Z_c(G) = G$, and in this case, for every $i = 0, 1, \ldots, c$

$$G^{c-i} \leq Z_i(G). \tag{8.3.7.1}$$

Proof. We prove this by induction on the minimal c such that either $G^c = \{1\}$ or $Z_c(G) = G$. When $c = 1$, either conditions $G^1 = \{1\}$ and $Z_1(G) = G$ is equivalent to the condition that G is abelian. The statement is clear.

Now suppose the theorem has been proved for smaller c, and we treat the case when either $G^c = \{1\}$ or $Z_c(G) = G$ for this $c \in \mathbb{N}$. Let $\pi : G \to G/Z(G) =: \overline{G}$. For a subgroup H of G, denote its image under π by $\overline{H} \subseteq \overline{G}$.

We first show that inductive hypothesis may be applied to $\overline{G}$ (whenever one of the condition of the theorem holds). In fact we will prove

$$\begin{tikzcd}
G^c = \{1\} \arrow[d, Leftrightarrow, "(1)"] & & Z_c(G) = G \arrow[d, Leftrightarrow, "(2)"] \\
(\overline{G})^{c-1} = \{1\} \arrow[rr, Leftrightarrow, "\text{inductive hypothesis}"] & & Z_{c-1}(\overline{G}) = \overline{G}.
\end{tikzcd}$$

Indeed, for (1), $G^c = \{1\}$ is equivalent to $[G, G^{c-1}] = \{1\}$, i.e. all elements in G^{c-1} commutes with all elements of G. So this is further equivalent to $G^{c-1} \leq Z(G)$. As the construction of lower central series is compatible with passing to quotients, this is further equivalent to $(\overline{G})^{c-1} = \{1\}$.

For statement (2), we simply note that the construction of upper central series implies inductively that $Z_i(\overline{G}) = \overline{Z_{i-1}(G)}$. More precisely, we have the following picture that explains this.

$$\begin{tikzcd}
G \arrow[d, "\pi_1"] \arrow[dd, bend right=60, "\pi_2"'] & \supseteq & Z_1(G) \arrow[rr, hook] & & Z_2(G) \arrow[d, "\pi_1"] \arrow[rr, hook] & & Z_3(G) \arrow[d, "\pi_1"] \\
G/Z_1(G) \arrow[rr, equal] & & \overline{G} \arrow[d, "\bar{\pi}_1"] & \supseteq & Z_1(\overline{G}) \arrow[rr, hook] & & Z_2(\overline{G}) \arrow[d, "\bar{\pi}_1"] \\
G/Z_2(G) \arrow[rr, equal] & & \overline{G}/Z_1(\overline{G}) \arrow[rr, equal] & & \overline{\overline{G}} & \supseteq & Z\big(\overline{G}/Z_1(\overline{G})\big).
\end{tikzcd}$$

By definition $Z_2(G)$ is the pullback of $Z_1(\overline{G})$ along π_1, namely $Z_2(G) = \pi_1^{-1}(Z_1(\overline{G}))$. After this, we consider $G/Z_2(G) \cong \overline{G}/Z_1(\overline{G})$. Inside of this, we will pullback $Z(G/Z_2(G)) =$

$Z(\overline{G}/Z_1(\overline{G}))$. The pullback to $\overline{G}$ gives $Z_2(\overline{G}) = \bar{\pi}_1^{-1}\big(Z(\overline{G}/Z_1(\overline{G}))\big)$. This implies that

$$Z_3(G) = \pi_2^{-1}\big(Z(\overline{G}/Z_1(\overline{G}))\big) = \pi_1^{-1}\bar{\pi}_1^{-1}\big(Z(\overline{G}/Z_1(\overline{G}))\big) = \pi_1^{-1}\big(Z_2(\overline{G})\big).$$

We may continue this way, and (2) is a special case of this results.

Finally, we prove $G^{c-i} \leq Z_i(G)$ from the inductive hypothesis $\overline{G}^{c-i} \leq Z_{i-1}(G)$. Indeed,

$$G^{c-i} \leq \pi^{-1}(\overline{G}^{c-i}) \leq \pi^{-1}(Z_{i-1}(\overline{G})) = Z_i(G). \tag{8.3.7.2}$$

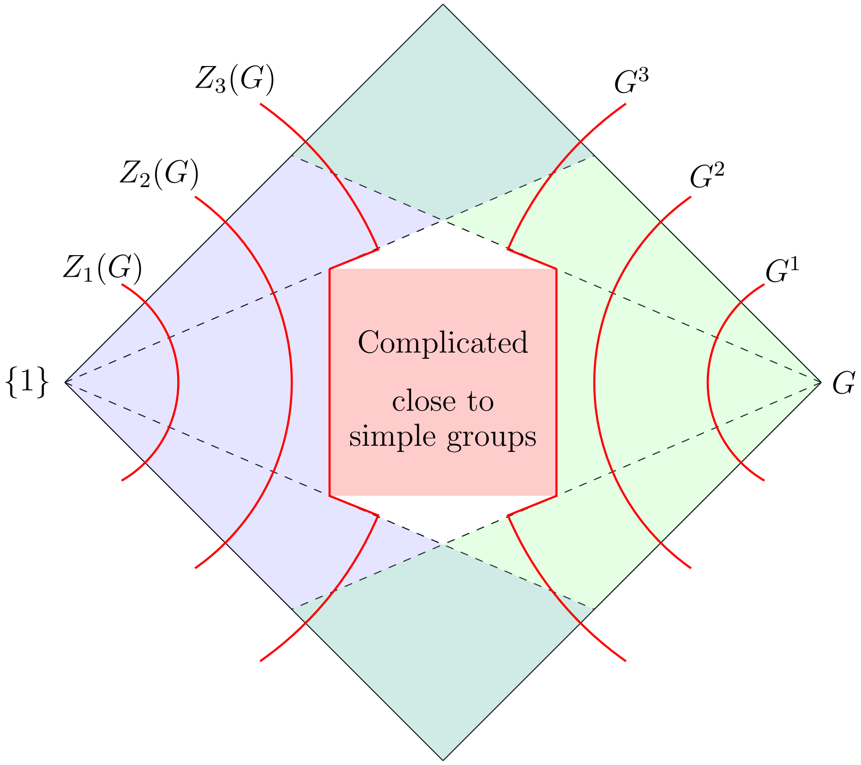
□

Remark 8.3.8. It is not true that $Z_i(G) \leq G^{c-i-1}$; this is an error in Dummit–Foote's book on page 194 Theorem 8. The essential reason is that $G^{c-i} \leq \pi^{-1}(\overline{G}^{c-i})$ in (8.3.7.2) need not be an equality, i.e. it is not true in general that the preimage of a commutator is a commutator.

Remark 8.3.9. One way to philosophically understand the lower or upper central series is that:

(1) abelian groups are easy to understand.
(2) If H is a nonabelian simple finite group, then $[H, H] = H$ and $Z(H) = 1$ (because H has no nontrivial normal subgroups). So lower and upper central series is trivial for simple finite groups.

The following visualize the lower and upper central series as follows.



Coming from the upper central series, we will never reach the Jordan–Hölder factors that "look like" simple groups. This is indicated on the left. "Dually", coming from the lower central series, the subgroup G^i will always contain the Jordan–Hölder factors taht "look like" simple groups. This filtration "shrinks" the group from the right of the picture.

8.4. **Structure theorem of nilpotent groups.** In fact, nilpotent groups have a very nice structure theorem.

Proposition 8.4.1. *All p-groups are nilpotent.*

Proof. This is because for every p-group P, $Z(P)$ is nontrivial by Proposition 6.3.4. □

Proposition 8.4.2. *Let P be a p-group.*

1. *We have $Z(P) \neq \{1\}$ (proved earlier).*
2. *If $H \trianglelefteq P$ is a* nontrivial *normal subgroup, then $H \cap Z(P) \neq \{1\}$.*
3. *If $H \lneqq P$, then $H \lneqq N_P(H)$.*

Proof. (1) is proved in Proposition 6.3.4.

(2) Consider the conjugation action of P on H:

$$P \overset{\mathrm{Ad}}{\curvearrowright} H, \qquad \mathrm{Ad}_p(h) = php^{-1}.$$

For this action, we write H as the disjoint union of orbits:

$$H = \coprod_i P/\mathrm{Stab}_P(a_i),$$

for representatives $a_1, a_2, \ldots, a_r$ of orbits.

We note that $\mathrm{Stab}_P(a_i) = P$ if and only if $a_i \in Z(P)$, namely $a_i \in Z(P) \cap H$. So we have the following

$$0 \equiv |H| = \sum_i |P| / \big|\mathrm{Stab}_P(a_i)\big| \equiv \big|Z(P) \cap H\big| \pmod p.$$

This implies that $Z(P) \cap H \neq \{1\}$.

(3) We use induction on $|P|$. There are two cases:

Case 1 If $Z(P) \nsubseteq H$, yet $Z(P) \subset N_P(H)$. So $H \lneqq N_P(H)$.

Case 2 If $Z(P) \subseteq H$, consider $\overline{H} = H/Z(P)$ and $\overline{P} = P/Z(P)$. By inductive hypothesis,

$$\overline{H} \lneqq N_{\overline{P}}(\overline{H}) \;\Rightarrow\; H \lneqq N_P(H).$$

□

Corollary 8.4.3. *If P is a p-group and if $H < P$ has index p, then H is a normal subgroup.*

Theorem 8.4.4 (Classification theorem for nilpotent groups)**.** *Let G be a finite group of order $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$ and $P_i \in \mathrm{Syl}_{p_i}(G)$. The following are equivalent.*

1. *G is nilpotent.*
2. *If $H \lneqq G$, then $H \lneqq N_G(H)$.*
3. *All Sylow subgroups P_i are normal.*
4. *$G \cong P_1 \times P_2 \times \cdots \times P_r$.*

Proof. (3) ⇒ (4) By criterion of direct products:

$$P_1P_2 = P_1 \times P_2, \quad P_1P_2P_3 = P_1P_2 \times P_3 = P_1 \times P_2 \times P_3, \;\ldots$$

(4) ⇒ (1) as each P_i is nilpotent.

(2) ⇒ (3) Recall that for each P_i, $N_G(N_G(P_i)) = N_G(P_i)$ by Corollary 7.2.6. So $N_G(P_i) = G$, which implies P_i is normal.

(1) ⇒ (2) Using the same argument as in the theorem above, noting that G is nilpotent ⇒ $G/Z(G)$ is nilpotent. □

8.5. **Schur–Zassenhaus theorem.** One may consult online resources on this topic, such as

`https://kconrad.math.uconn.edu/blurbs/grouptheory/schurzass.pdf`

We have learned in the Hölder program that, if N is a normal subgroup of G, then the study of G, philosophically, may be reduced to the study of N and G/N. The Schur–Zassenhaus theorem concerns how to "reconstruct" G from N and G/N. This is in general difficult. For example, if N and G/N are both isomorphic to $\mathbf{Z}_p$, then G can be either $\mathbf{Z}_{p^2}$ or $\mathbf{Z}_p \times \mathbf{Z}_p$. The case of $\mathbf{Z}_p \times \mathbf{Z}_p$ is considered a little "better" because the quotient map $\pi : G \to G/N$ admits a "partial inverse" or a section $i : G/N \to G$ such that $\pi \circ i = \mathrm{id}_{G/N}$.

What Schur and Zassenhaus proved is that when $|N|$ and $|G/N|$ are relatively prime, then there is always such a section i.

Theorem 8.5.1 (Schur–Zassenhaus)**.** *Let G be a finite group and $N \trianglelefteq G$ a normal subgroup such that $|N|$ and $|G/N|$ are relatively prime. Let $\pi : G \twoheadrightarrow G/N$ be the natural surjective homomorphism.*

(1) There exists a subgroup $H \leq G$ such that

$$\pi|_H : H \to G/N$$

is an isomorphism. (Such H's are called **complements** *of N in G.) As a corollary, G is isomorphic to a semidirect product $N \rtimes (G/N)$.*

(2) When either N or G/N is solvable, all complement subgroups H in (1) are conjugate to each other.

Remark 8.5.2. Using the difficult theorem of Feit–Thompson (Remark 3.3.5), the condition in (2) is always satisfied: because either N or G/N must have odd cardinality.

The proof of Schur–Zassenhaus theorem is divided into two parts:

- Step I: Prove Schur–Zassenhaus theorem in the case when N is an abelian group.
- Step II: Reduce to the case when N is an abelian group.

The proof of Step I makes use of the so-called group cohomology which is beyond the scope of abstract algebra. Essentially, constructing G from N and G/N is governed by a cohomological class in $H^2(G/N, N)$. Since $|N|$ and $|G/N|$ are relatively prime, $H^2(G/N, N)$ is trivial. This can be directly interpreted as the existence of a complement group. The fact that all complement groups are conjugate follows from the vanishing of $H^1(G/N, N)$. (Although we do not cover these in our honors algebra course, group cohomology is a very foundamental mathematical objects in algebra with many applications.)

In this supplementary material, we focus on the proof of Step II (of the existence part). We start with a lemma.

Lemma 8.5.3. *If $N \trianglelefteq G$ and $P \in \mathrm{Syl}_p(N)$, then $G = N \cdot N_G(P)$. In particular, if $P \trianglelefteq N$, then $P \trianglelefteq G$.*

Proof. Pick $g \in G$, $gPg^{-1} \trianglelefteq gNg^{-1} = N$. So gPg^{-1} is a Sylow p-subgroup of N. By 2nd Sylow Theorem, there exists $n \in N$ such that $gPg^{-1} = nPn^{-1}$. It then follows that $n^{-1}gPg^{-1}n = P$, i.e. $n^{-1}g \in N_G(P)$. So $g \in nN_G(P) \subseteq N \cdot N_G(P)$.

To see the second statement under the condition $P \trianglelefteq N$, we note that in this case $N \subseteq N_G(P)$, so $G = N \cdot N_G(P) = N_G(P)$, i.e. $P \trianglelefteq G$. □

Proof of Step II of existence in Schur–Zassenhaus theorem. By induction on the order of the group G, we can assume that Theorem 8.5.1(1) holds for smaller groups.

(a) If N strictly contains a nontrivial subgroup N' that is normal in G, let $\overline{G} := G/N'$ and $\overline{N} := N/N'$. Then $G/N \cong \overline{G}/\overline{N}$. Applying inductive hypothesis to $\overline{G}$, there exists a complement $\overline{M} \leq \overline{G}$ (that is isomorphic to $\overline{G}/\overline{N} \cong G/N$).

If $\pi : G \to G/N'$ denote the projection, let $M := \pi^{-1}(\overline{M})$, and then N' is a normal subgroup of M such that $M/N' \cong \overline{M}$. Applying inductive hypothesis again to M shows that N' has a complement H in M, namely a subgroup $H \leq M \leq G$ such that

$$H \to \overline{H} \to G/N$$

is an isomorphism.

(b) Suppose now that N is a minimal normal subgroup of G.

Let P be a Sylow subgroup of N (for some prime number p). Lemma 8.5.3 implies that $G = N \cdot N_G(P)$. By 2nd Isomorphism Theorem,

$$G/N \cong N_G(P)\big/(N_G(P) \cap N).$$

If $N_G(P) \neq G$, the inductive hypothesis may be applied to $N_G(P)$ to obtain a complement of $N_G(P) \cap N$ in $N_G(P)$, which may be also served as a complement of N in G.

If $N_G(P) = G$, then P is a normal subgroup of G. By minimality of N, $N = P$. By normality of N, $N = P$ is the unique Sylow p-subgroup of G. Since $Z(P)$ is a characteristic subgroup of P, $Z(P)$ is normal in G. By minimality of N again, $N = Z(P)$ is an abelian group. This reduces to the case where we can handle by Galois cohomology. □

9. Rings, ideals, quotients rings

9.1. Rings.

Definition 9.1.1. A **ring** R is a set together with two binary operations $+$ and $\cdot$, satisfying

1. $(R, +)$ is an abelian group under "addition" (with 0 as the additive unit);
2. the "multiplication" $\cdot$ is associative, i.e. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for any $a, b, c \in R$;
3. the *distributive law* holds in R, i.e., for all $a, b, c \in R$,
$$(a+b) \cdot c = a \cdot c + b \cdot c \quad \text{and} \quad a \cdot (b+c) = a \cdot b + a \cdot c;$$
4. R is *unital*, i.e. there exists an element $1 \in R$ such that $1 \neq 0$ and that
$$1 \cdot a = a \cdot 1 = a, \quad \text{for all } a \in R.$$

In this course, all rings are assumed to be unital and $1 \neq 0$, i.e. condition (4) above holds.

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

Definition 9.1.2. A ring is called a **division ring** or a **skew field** if every nonzero element $a \in R$ has a multiplicative inverse.

A commutative division ring is called a **field**.

Example 9.1.3. (1) $(\mathbb{Z}, +, \cdot)$ and $(\mathbf{Z}_n, +, \cdot)$ are rings.

(2) $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields.

(3) $\mathbb{Z}[\frac{1}{N}] = \big\{\frac{a}{N^r} \,\big|\, a \in \mathbb{Z}, r \in \mathbb{Z}_{\geq 0}\big\}$ is a subring of $\mathbb{Q}$.

(4) If R is a ring, then $R[x] = \Big\{\sum\limits_{n \geq 0} a_n x^n \,\Big|\, a_n \in R\Big\}$ is a ring, called the **polynomial ring over** R. More generally, we may define $R[x_1, \dots, x_n]$ similarly as the ring of multivariable polynomial rings with coefficients in R.

In this construction, we often require R to be commutative.

(5) If R is a ring, then the set of all $n \times n$ matrices $\mathrm{Mat}_{n \times n}(R)$ in R is a ring.

(6) $\mathbb{H} := \big\{a + bi + ci + dk \,\big|\, a, b, c, d \in \mathbb{R}\big\}$ is called the ring of **Hamilton quaternions**. The multiplication is given by the rules:
$$i^2 = j^2 = k^2 = -1,\ ij = -ji = k,\ jk = -kj = i,\ ki = -ik = j.$$
There are additional structures, for $z = a + bi + cj + dk \in \mathbb{H}$,

- Conjugations: $\bar{z} = \overline{a + bi + cj + dk} := a - bi - cj - dk$;
- $\overline{z \cdot w} = \bar{w} \cdot \bar{z}$;
- Norm map $\mathrm{Nm}(z) := z\bar{z} = a^2 + b^2 + c^2 + d^2 \in \mathbb{R}_{\geq 0}$.

It can be seen that if $z \neq 0$, then $z^{-1} = \bar{z}/\mathrm{Nm}(z)$ is a multiplicative inverse. So $\mathbb{H}$ is a division ring. (See extended reasons for more discussion.)

(7) (Group rings) Let R be a commutative and G a group. Define the associated **group ring** of G over R to be
$$R[G] := \Big\{\text{finite sums} \sum_{g \in G} a_g g \,\Big|\, a_g \in R\Big\}.$$
The multiplication is given by
$$\Big(\sum_{g \in G} a_g g\Big)\Big(\sum_{h \in G} b_h h\Big) = \sum_{g, h \in G} a_g b_h gh.$$

The multiplicative unit in $R[G]$ is $1 \cdot e_G$ (where e_G is the unit in G).

For example, when $G = \mathbf{Z}_n = \langle \sigma | \sigma^n = 1 \rangle$, we have

$$R[G] = \big\{ a_0 + a_1\sigma + \cdots + a_{n-1}\sigma^{n-1} \,\big|\, a_i \in R \big\},$$

subject to the rule that $\sigma^n = 1$.

For another example, $G = \mathbb{Z} = \langle \sigma \rangle$, then

$$R[G] = R[x^{\pm 1}] = \Big\{ \text{finite sum } \sum_{n \in \mathbb{Z}} a_n x^n \,\Big|\, a_n \in R \Big\}.$$

Definition 9.1.4. Let R and S be rings.

(1) A **ring homomorphism** is a map $\phi : R \to S$ satisfying
   (a) $\phi(a+b) = \phi(a) + \phi(b)$ for all $a, b \in R$ (which in particular implies that $\phi(0) = 0$),
   (b) $\phi(ab) = \phi(a)\phi(b)$ for all $a, b \in R$,
   (c) $\phi(1) = 1$.

(2) The **kernel** of ϕ is $\ker \phi = \phi^{-1}(0)$. In particular, ϕ is injective if and only if $\ker \phi = \{0\}$.

(3) A homomorphism ϕ is called an **isomorphism** if it is bijective.

Remark 9.1.5. In some books, a ring is not assumed to contain 1 and a ring homomorphism needs not to send 1 to 1. We impose both conditions, as this is the case we almost always encounter in the future.

In particular, if R_1 and R_2 are two rings, the natural map $R_1 \to R_1 \times R_2$ given by $a \mapsto (a, 0)$ is NOT a homomorphism, because it does not send 1 to 1. This is actually quite reasonable; see our later discussion on the prime spectrum of a commutative ring.

Example 9.1.6. (1) $\phi : \mathbb{Z} \to \mathbf{Z}_n$ sending a to $a \bmod n$ is a homomorphism.

(2) $\phi : R \to S$ sending all elements of R to $0 \in S$ is not a homomorphism (under our definition).

(3) $\phi : \mathbb{Z} \to \mathbb{Z}$ sending $\phi(x) = nx$ is not a homomorphism unless $n = 1$.

(4) If R is a *commutative* ring, then for any $r \in R$, there is a natural **evaluation** homomorphism:

$$\begin{aligned} \phi_r : R[x] &\longrightarrow R \\ f(x) &\longmapsto f(r). \end{aligned}$$

Note that it is crucial to assume R to be commutative here!

Definition 9.1.7. Let R be a ring.

(1) A nonzero element $a \in R$ is called a **zero-divisor** if there exists a nonzero element $b \in R$ such that either $ab = 0$ or $ba = 0$.

(2) $u \in R$ is called a **unit** in R if there exists $v \in R$ such that

$$uv = vu = 1.$$

The set of units in R is $R^\times$. They form a group under multiplication.

A *commutative* ring R containing no zero-divisor is called an **integral domain**.

Remark 9.1.8. A key property of an integral domain is the following *cancellation law*: if $a \in R$ and $a \neq 0$, then for any $b, c \in R$,

$$a \cdot b = a \cdot c \quad \Rightarrow \quad b = c.$$

Example 9.1.9. (1) $\mathbf{Z}_n^\times = \big\{a \bmod n \,\big|\, \gcd(a,n) = 1\big\}$. The zero divisors in $\mathbf{Z}_n$ are $\big\{a \bmod n \neq 0 \bmod n \,\big|\, \gcd(a,n) \neq 1\big\}$.

(2) If R is an integral domain, then so is $R[x]$.

This is because if one has two nonzero polynomials $f(x) = a_mx^m + \cdots + a_0$ and $g(x) = b_nx^n + \cdots + b_0$ (with $a_m \neq 0$ and $b_n \neq 0$), then $f(x)g(x)$ has leading term $a_mb_nx^{m+n}$. But $a_mb_n \neq 0$, so $f(x)g(x) \neq 0$. This shows that $R[x]$ is an integral domain.

Lemma 9.1.10. *A finite integral domain R is a field.*

Proof. For any nonzero element $a \in R$, we need to find its inverse. Consider the following homomorphism of additive groups

$$\begin{aligned} \phi_a : (R,+) &\longrightarrow (R,+) \\ x &\longmapsto ax \end{aligned}$$

Then $\ker \phi_a = \{x \in R \,|\, ax = 0\} = \{0\}$ as R is an integral domain. Thus ϕ_a is injective, and hence an isomorphism by counting the number of elements.

In particular, $a^{-1} := \phi_a^{-1}(1)$ is a multiplicative inverse of a. □

Definition 9.1.11. For an integral domain R, we define its **fraction field** or the **quotient field** to be

$$\mathrm{Frac}(R) := \Big\{(a,b) \in R \times \big(R \backslash \{0\}\big)\Big\} \Big/ \big((a,b) \sim (c,d) \text{ if and only if } ad = bc\big).$$

In particular, R is a field.

Example 9.1.12. (1) $\mathrm{Frac}(\mathbb{Z}) = \mathbb{Q}$.

(2) For a field k, $\mathrm{Frac}(k[x]) = k(x)$ the field of rational functions in x (with coefficients in k).

9.2. **Ideals.** To develop the story for rings in a parallel way to that of groups, we now introduce the analogue of normal groups in the theory of rings.

Definition 9.2.1. A subset $I \subseteq R$ is called a **left ideal** if

(1) for any $a, b \in I$, $a - b \in I$ (so that I is a subgroup of $(R,+)$);

(2) for any $a \in I$ and $x \in R$, we have $xa \in I$.

We say that I is a **right ideal** if it satisfies the above conditions with (2) replaced by $ax \in I$.

We say that I is an **ideal** (or a **two-sided ideal**) if it is a left ideal and a right ideal as the same time.

We say that I is a **proper ideal** if $I \neq R$.

For commutative rings, there is no difference between left, right, or two-sided ideals. So when R is not known to be commutative, we will always say two-sided ideals, but if R is commutative, we will simply say ideals.

Remark 9.2.2. An ideal of a ring is (usually) not a ring, because $1 \notin I$. ($1 \in I$ implies that $I = R$.)

We represent ideals using the following notation

Notation 9.2.3. Let R be a commutative ring, and let $a_1, \dots, a_s \in R$. Define

$$\big(a_1, \dots, a_s\big) = \Big\{ \sum_{i=1}^{s} x_i a_i \,\Big|\, x_i \in R \Big\} \subseteq R.$$

(This definition also works for infinite sets $\{a_i | i \in J\}$ when we allow only finite sums.)

We call this ideal the **ideal generated by** $a_1, \dots, a_s$. It is the minimal ideal that contains all of $a_1, \dots, a_s$.

Example 9.2.4. In $R = \mathbb{Z}$,

$$(4, 6) = \{4x + 6y \,|\, x, y \in \mathbb{Z}\} = 2\mathbb{Z} = (2).$$

In general, $(a_1, \dots, a_s) = \big(\gcd(a_1, \dots, a_s)\big)$.

Definition 9.2.5. Let R be a ring and I a two-sided ideal such that $I \neq R$. We define the **quotient ring** $R/I := \{x + I \,|\, x \in R\}$ (quotient as an additive group) with operations:

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

We check that the multiplication is well-defined: if $x' = x + a$ and $y' = y + b$ with $a, b \in I$, then

$$x'y' + I = (x + a)(y + b) + I = xy + \underbrace{xb + ay + ab}_{\text{each term} \in I} + I = xy + I.$$

There is a natural surjective quotient homomorphism

$$\begin{tikzcd} \pi : R \arrow[r, two heads] & R/I \\ x \arrow[r, mapsto] & x + I =: \bar{x} \end{tikzcd}$$

with $\ker \pi = I$.

This following is the analogue of isomorphism theorems for rings.

Theorem 9.2.6 (Isomorphism Theorems). (1) *If $\phi : R \to S$ is a ring homomorphism, then* $\ker \phi$ *is a two-sided ideal and $\phi(R)$ is a subring of S.*

Moreover, ϕ induces an isomorphism

$$\begin{tikzcd} R/\ker \phi \arrow[r, "\cong"] & \phi(R) \\ x + \ker \phi \arrow[r, mapsto] & \phi(x) \end{tikzcd}$$

(2) *Let $I \subseteq J$ be proper ideals of R, then $J/I \subseteq R/I$ is a proper ideal and*

$$(R/I)\big/(J/I) \cong R/J.$$

(3) *Let I be a proper ideal of R. Then there is a 1-1 correspondence*

$$\begin{tikzcd} \big\{\textit{left/right/two-sided ideals } J \textit{ containing } I\big\} \arrow[r, leftrightarrow] & \big\{\textit{left/right/two-sided ideals } \bar{J} \textit{ of } R/I\big\} \\ J \arrow[r, mapsto] & J/I \\ \pi^{-1}(\bar{J}) & \bar{J} \arrow[l, mapsto] \end{tikzcd}$$

preserving inclusion orders, sums, intersections, and quotients. (Sum of ideals will soon be defined.)

Example 9.2.7. (1) For the homomorphism $\phi : \mathbb{Z} \to \mathbf{Z}_n$ given by modulo n, $\ker\phi = n\mathbb{Z} = (n)$. So $\mathbf{Z}_n \cong \mathbb{Z}/n\mathbb{Z}$.

(2) For a commutative ring R and $a \in R$, we have an evaluation homomorphism

$$\begin{aligned} \phi_a : R[x] &\longrightarrow R \\ f(x) &\longmapsto f(a). \end{aligned}$$

The kernel $\ker\phi_a = \big\{f(x) \in R[x] \,\big|\, f(a) = 0\big\} = (x-a)$. (This is because one can always write every $f(x) \in R[x]$ as $f(x) = g(x)(x-a) + f(a)$. So $R[x]/(x-a) \cong R$.

(3) Let R be a ring and G a group, there is a natural homomorphism from the group ring $R[G]$:

$$\begin{aligned} \phi : R[G] &\longrightarrow R \\ \sum_{g \in G} a_g[g] &\longmapsto \sum_{g\in G} a_g. \end{aligned}$$

The kernel $\ker\phi = (g-1; g \in G)$ is called the **augmentation ideal** of $R[G]$.

(4) For $R = R_1 \times R_2$ a direct product of rings, both $R_1 \times \{0\}$ and $\{0\} \times R_2$ are ideals. They cay be alternatively written as

$$R_1 \times \{0\} = ((1,0)) \quad \text{and} \quad \{0\} \times R_2 = ((0,1)).$$

Caveat: the map

$$\begin{aligned} R_1 &\longrightarrow R_1 \times R_2 \\ a &\longmapsto (a,0) \end{aligned}$$

does not take 1_{R_1} to $1_{R_1 \times R_2}$; so it is NOT a homomorphism in our convention.

We have the following operations of ideals.

Definition 9.2.8. Let I and J be two-sided ideals of a ring R.

(1) Define the **sum of ideals** to be

$$I + J = \big\{a + b \,\big|\, a \in I,\, b \in J\big\}.$$

(2) Define the **product of ideals** to be

$$IJ = \big\{\text{finite sums of elements } ab \text{ for } a \in I,\, b \in J\big\}.$$

Caveat 9.2.9. In general, it is not true that all elements of IJ can be written as a pure product ab for $a \in I$ and $b \in J$. For example, $R = \mathbb{Z}[x]$ and $I = (2, x) = \big\{f(x) \in \mathbb{Z}[x] \,\big|\, f(0) = 2\big\}$. Then $x^2 + 4 \in I^2$ yet it cannot be written in the form of ab with $a, b \in I$.

Remark 9.2.10. By the property of ideals, $IJ \subseteq I$ and $IJ \subseteq J$, so

$$IJ \subseteq I \cap J.$$

Example 9.2.11. If R is a commutative ring and $I = (a_1, \dots, a_s)$ and $J = (b_1, \dots, b_t)$, then

$$I + J = (a_1, \dots, a_s, b_1, \dots, b_t), \qquad IJ = (a_1b_1, \dots, a_1b_t, \dots, a_ib_j, \dots, a_sb_t).$$

Remark 9.2.12. It is important to explain the practical meaning of taking quotient rings: it is to **imposing relations among generators**. We explain this through an example: for k a field, we show that

$$k[x,y,z]\big/(x-y^2,\, y-z^3) \cong k[z].$$

Indeed, for example, the element x^2y can be written as

$$\begin{aligned} x^2y\ &= (x-y^2+y^2)^2y = (x-y^2)\cdot * + y^4y = (x-y^2)\cdot * + (y-z^3+z^3)^5 \\ &= (x-y^2)\cdot * + (y-z^3)\cdot * + z^{15}. \end{aligned}$$

So x^2y is equivalent to z^{15} in the quotient.

For another example, consider a homomorphism

$$\begin{aligned} \phi_i : \mathbb{R}[x] &\longrightarrow \mathbb{C} \\ f(x) &\longmapsto f(i) \end{aligned}$$

The kernel $\ker \phi_i = (x^2+1)$. So

$$\mathbb{R}[x]/(x^2+1) \cong \mathbb{C}.$$

(namely, we are imposing the relation $x^2+1=0$ in the ring.) We also point out that this is the prototype for field extension later, describing $\mathbb{C}$ in terms $\mathbb{R}$.

Extended readings after Section 11

9.3. **Quaternions over $\mathbb{Q}$.** In fact, in the constructions of Hamilton quaternions, we do not really need it to have coefficients in $\mathbb{R}$. In fact, for nonzero numbers $A, B \in \mathbb{Q}$, we may define a quaternion ring over $\mathbb{Q}$:

$$D_{A,B} := \big\{ a + bi + cj + dij \,\big|\, a, b, c, d \in \mathbb{Q} \big\}$$

where the multiplications are $\mathbb{Q}$-linear and are governed by

$$i^2 = A, \quad j^2 = B, \quad ij = -ji.$$

When $A = B = -1$ and if we change the coefficients from $\mathbb{Q}$ to $\mathbb{R}$, then we recover the Hamilton quaternions.

It is an interesting fact (which is also important in number theory) that such $D_{A,B}$ is either isomorphic to the matrix ring $\mathrm{Mat}_{2\times 2}(\mathbb{Q})$ or is a division ring. (For example, if both A and B are negative then $D_{A,B}$ is a division ring for the "same" reason as in for $\mathbb{H}$. Yet for each p, if A is exactly divisible by p and B is an integer whose reduction modulo p is not a square, then $D_{A,B}$ is a division ring "for reasons at p".)

10. Chinese remainder theorem, Euclidean domains, and PIDs

10.1. **Chinese remainder theorem.**

Recall that the classical Chinese remainder theorem can be restated as follows: if $n_1, \dots, n_r$ are pair-wise coprime integers, then

$$\mathbb{Z} \longrightarrow \mathbb{Z}/n_1\mathbb{Z} \times \cdots \mathbb{Z}/n_r\mathbb{Z}$$

is surjective and its kernel is $n_1\mathbb{Z} \cap \cdots \cap n_r\mathbb{Z} = n_1 \cdots n_r\mathbb{Z}$.

Definition 10.1.1. Let R be a two commutative ring. We say two ideals I and J of R are **comaximal** if $I + J = R$, i.e. $1 \in R$ can be written as $1 = a + b$ with $a \in I$ and $b \in J$.

(Note that in the case $R = \mathbb{Z}$, $m, n \in \mathbb{Z}$ are coprime if and only if $(m)+(n) = (\gcd(m,n)) = (1)$.)

Theorem 10.1.2. *Let $I_1, \dots, I_k$ be ideals of a commutative ring R. Then the natural map*

$$\begin{aligned} \phi : R &\longrightarrow R/I_1 \times \cdots \times R/I_k \\ x &\longmapsto (x \bmod I_1, \dots, x \bmod I_k) \end{aligned}$$

is a ring homomorphism with kernel $I_1 \cap \cdots \cap I_k$.

If $I_1, \dots, I_k$ are pairwise comaximal, then

1. *ϕ is surjective, and*
2. *$I_1 \cap \cdots \cap I_k = I_1 \cdots I_k$.*

In particular, this implies that

$$\phi : R/I_1 \cdots I_k \cong R/I_1 \cap \cdots \cap I_k \xrightarrow{\simeq} R/I_1 \times \cdots \times R/I_k.$$

Proof. The first claim on ϕ being a homomorphism with kernel $I_1 \cap \cdots \cap I_k$ is clear. We now prove (1) and (2).

We first assume that $k = 2$. As $I_1I_2 \subseteq I_1$ and $I_1I_2 \subseteq I_2$, we have $I_1I_2 \subseteq I_1 \cap I_2$. Now, if $R = I_1 + I_2$, we may write $1 = a_1 + a_2$ with $a_1 \in I_1$ and $a_2 \in I_2$. Then for $b \in I_1 \cap I_2$,

$$b = \underbrace{ba_1}_{\text{in } I_2I_1} + \underbrace{ba_2}_{\text{in } I_1I_2} \in I_1I_2.$$

This implies that $I_1I_2 = I_1 \cap I_2$.

To see that ϕ is surjective in this case, we note that

$$\begin{aligned} \phi(a_1) &= \big(a_1 \bmod I_1,\ a_1 = 1 - a_2 \bmod I_2\big) = (0, 1); \\ \phi(a_2) &= \big(a_2 = 1 - a_1 \bmod I_1,\ a_2 \bmod I_2\big) = (1, 0); \end{aligned}$$

Thus, for any $(x_1 \bmod I_1,\ x_2 \bmod I_2) \in A/I_1 \times A/I_2$, it is $\phi(a_1x_2 + a_2 + x_1)$.

In general, we use induction to show

$$\phi : R \longrightarrow R/I_1 \times R/I_2 \cdots I_k \twoheadrightarrow R/I_1 \times \cdots \times R/I_k.$$

For this, we need to check I_1 and $I_2 \cdots I_k$ are comaximal, i.e. $I_1 + I_2 \cdots I_k = R$. This is because for each $i = 2, \dots, k$, $1 = a_i + b_i$ for $a_i \in I_1$ and $b_i \in I_i$. Thus

$$1 = (a_2 + b_2) \cdots (a_k + b_k) = \underbrace{a_2 \cdots a_k + \text{product with some } a_i}_{\text{in } I_1} + \underbrace{b_1 \cdots b_k}_{\text{in } I_2 \cdots I_k}.$$

□

10.2. A digression in logic.

Definition 10.2.1. A **partial order** on a nonempty set A is a relation $\preceq$ on A satisfying for all $x, y, z \in A$,

1. (reflexive) $x \preceq x$;
2. (antisymmetric) if $x \preceq y$ and $y \preceq x$, then $x = y$;
3. (transitive) if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

(Some times we say A is a **poset**.)

A **chain** is a subset $B \subseteq A$ where for any $x, y \in B$, either $x \preceq y$ or $y \preceq x$.

Axiom 10.2.2 (Zorn's Lemma)**.** If A is a partially ordered set in which every chain B has an upper bound, i.e. an element $m \in A$ such that $m \succeq b$ for every $b \in B$, then A has a maximal element x, i.e. an element such that no $y \succ x$.

Zorn's Lemma is independent of the Zermelo–Fraenkel axiom system and is equivalent to the Axiom of Choice and the Well-ordering Principle. See for example Dummit–Foote's Appendix A.2 or Munkres' Topology for more discussion.

In this lecture, we assume that Zorn's lemma holds.

10.3. Maximal ideals.

Definition 10.3.1. If R is a ring, a (two-sided) ideal $\mathfrak{m} \subseteq R$ is called **maximal** if $\mathfrak{m} \neq R$ and the only (two-sided) ideals containing $\mathfrak{m}$ are $\mathfrak{m}$ and R.

The existence of such maximal ideals relies on the Zorn's lemma.

Proposition 10.3.2. *Every proper (two-sided) ideal $I \subsetneq R$ is contained in a maximal ideal of R.*

Proof. Put $\mathscr{S} := \{$proper ideals of R containing $I\}$. I claim that it is a partially ordered set for inclusion. For this, we need to check that every increasing chain $J_i \subseteq \cdots$ of ideals has an upper bound. Indeed, $J = \bigcup_{i \in S} I_i$ is an ideal; yet $1 \neq J$; so J is a proper ideal containing I.

So by Zorn's lemma, $\mathscr{S}$ admits a maximal element, namely, the maximal ideal needed. □

The following is an important criterion for maximal ideals.

Proposition 10.3.3. *Let R be a commutative ring. An ideal $\mathfrak{m} \subseteq R$ is maximal if and only if the quotient $R/\mathfrak{m}$ is a field.*

Proof. By lattice isomorphism theorem, $\mathfrak{m} \subseteq R$ if and only if $\bar{R} := R/\mathfrak{m}$ has only two ideals (0) and (1). We claim that the latter statement is equivalent to that $\bar{R}$ is a field.

If $\bar{R}$ is a field, then clearly it has only two ideals (0) and (1). Conversely, if $\bar{R}$ has only two ideals (0) and (1), then for any nonzero element $a \in \bar{R}$, the ideal $(a) \neq (0)$. Thus $(a) = (1)$, namely there exists $a' \in \bar{R}$ such that $aa' = 1$. This implies that $a \in \bar{R}^\times$. So $\bar{R}$ is a field. □

Remark 10.3.4. If R is non-commutative, then $R/\mathfrak{m}$ being a skew field implies that $\mathfrak{m}$ is maximal. But the converse is not correct. For example, $R = \mathrm{Mat}_{n \times n}(\mathbb{C})$ has only two two-sided ideals: (0) and R. Yet R is not a skew field.

Example 10.3.5.

1. When $R = \mathbb{Z}$, for each prime number p, $(p) = p\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$.

(2) For $R = \mathbb{Z}[x]$, $(p) = p\mathbb{Z}[x]$ is not a maximal ideal. But (p, x), or $(p, x+1)$, or $(p, f(x))$ for any polynomial irreducible modulo p, is a maximal ideal.

(3) For G a finite group and $R = \mathbb{C}[G]$ the group ring, the augmentation ideal $I_G = \big\langle [g] - 1 \,\big|\, g \in G \big\rangle$ is a maximal (two-sided) ideal, and we have $\mathbb{C}[G]/I_G \cong \mathbb{C}$.

10.4. **Prime ideals.** Assume from now on that R is commutative.

Definition 10.4.1. A proper ideal $\mathfrak{p} \subsetneq R$ is called a **prime ideal** if

$$\text{for any } a, b \in R, \quad ab \in \mathfrak{p} \Rightarrow a \in \mathfrak{p} \text{ or } b \in \mathfrak{p}.$$

This is equivalent to

$$a \notin \mathfrak{p} \text{ and } b \notin \mathfrak{p} \;\Rightarrow\; ab \notin \mathfrak{p}.$$

Example 10.4.2. For p a prime number, $p\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$ and $p\mathbb{Z}[x] \subset \mathbb{Z}[x]$ is also a prime ideal.

The following is an analogue of Proposition 10.3.3 for prime ideals. This is also quite useful in application.

Proposition 10.4.3. *An ideal $\mathfrak{p} \subset R$ is a prime ideal if and only if $R/\mathfrak{p}$ is an integral domain.*

Proof. Consider the natural quotient

$$\begin{aligned} \pi : R &\longrightarrow\!\!\!\!\!\rightarrow R/\mathfrak{p} \\ a &\longmapsto \bar{a}. \end{aligned}$$

If $R/\mathfrak{p}$ is an integral domain, then for $a, b \in R$ with $ab \in \mathfrak{p}$, we must have $\overline{ab} = 0$, or equivalently $\bar{a}\bar{b} = 0$. This means that either $\bar{a} = 0$ or $\bar{b} = 0$, i.e. either $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$.

Conversely, suppose that $R/\mathfrak{p}$ is not an integral domain, then there exists nonzero elements $\bar{a}, \bar{b} \in R/\mathfrak{p}$ such that $\bar{a}\bar{b} = 0$. This is equivalent to say that there exists $a, b \in R \backslash \mathfrak{p}$ such that $ab \in \mathfrak{p}$. Thus $\mathfrak{p}$ is not a prime ideal in this case. □

Corollary 10.4.4. *A maximal ideal is always a prime ideal.*

The concepts of maximal ideals and prime ideals are extremely important in the study of commutative algebra.

The following is an interesting property of prime ideals. The method of proof is typical in commutative algebra.

Proposition 10.4.5. (1) *Let $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ be ideals and let $\mathfrak{p}$ be a prime ideal containing $\bigcap_{i=1}^{n} \mathfrak{a}_i$. Then $\mathfrak{p} \supseteq \mathfrak{a}_i$ for some i. If $\mathfrak{p} = \bigcap \mathfrak{a}_i$, then $\mathfrak{p} = \mathfrak{a}_i$ for some i.*

(2) *Let $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ be prime ideals and let $\mathfrak{a}$ be an ideal contained in $\bigcup_{i=1}^{n} \mathfrak{p}_i$. Then $\mathfrak{a} \subseteq \mathfrak{p}_i$ for some i.*

Proof. (1) Suppose not. Then we may take elements $x_i \in \mathfrak{a}_i \backslash \mathfrak{p}$ for each i. Yet since $\mathfrak{p}$ is a prime ideal,

$$x_1 x_2 \cdots x_n \notin \mathfrak{p} \quad \text{but} \quad x_1 x_2 \cdots x_n \in \bigcap_{i=1}^{n} \mathfrak{a}_i.$$

Thus $\mathfrak{p} \supset \mathfrak{a}_i$ for some i.

If moreover $\mathfrak{p} = \bigcap_{j=1}^{n} \mathfrak{a}_j$, then $\mathfrak{p} \subseteq \mathfrak{a}_j$ for each j, and thus $\mathfrak{p} = \mathfrak{a}_i$ for the i above.

(2) We prove by induction on n that

$$\mathfrak{a} \not\subseteq \mathfrak{p}_i \text{ for } i = 1, \dots, n \implies \mathfrak{a} \not\subseteq \bigcup_{i=1}^{n} \mathfrak{p}_i. \tag{10.4.5.1}$$

The base case $n = 1$ is clear, and suppose that we have proved (10.4.5.1) for $n - 1$, and now we prove this for n. By inductive hypothesis, for each $i = 1, \dots, n$, we may find $x_i \in \mathfrak{a}$ such that

$$x_i \notin \mathfrak{p}_1 \cup \dots \cup \mathfrak{p}_{i-1} \cup \mathfrak{p}_{i+1} \cup \dots \cup \mathfrak{p}_n.$$

If for some x_i, $x_i \notin \mathfrak{p}_i$, we have already verified (10.4.5.1) using this element x_i. Now, it suffices to treat the case when $x_i \in \mathfrak{p}_i$ for every $i = 1, \dots, n$. Then consider the element

$$y = \sum_{i=1}^{n} x_1 \cdots x_{i-1} x_{i+1} \cdots x_n.$$

Clearly $y \in \mathfrak{a}$. We show that $y \notin \mathfrak{p}_i$ for every i. Indeed, all elements in the sum except the ith term $x_1 \cdots x_{i-1} x_{i+1} \cdots x_n$ belongs to $\mathfrak{p}_i$. But as $\mathfrak{p}_i$ is a prime ideal, this product $x_1 \cdots x_{i-1} x_{i+1} \cdots x_n$ does not belong to $\mathfrak{p}_i$. So $y \notin \mathfrak{p}_i$. This completes the inductive proof. □

10.5. **Moving ideals along homomorphism of rings.**

Notation 10.5.1. Let $f : R \to S$ be a homomorphism of commutative rings.

1. If $J \subseteq S$ is an ideal, then $f^{-1}(J)$ is an ideal, sometimes called the **contraction** of J.
2. If $I \subseteq R$ is an ideal of R, then $f(I)S$ is an ideal of S, sometimes called the **extension** of the ideal I. (Note that $f(I)$ need not to be an ideal of S.)

Proposition 10.5.2. *If $J \subseteq S$ is a prime ideal, then $f^{-1}(J)$ is a prime ideal of R.*

Proof. We consider the natural homomorphism

$$\varphi : R \xrightarrow{f} S \twoheadrightarrow S/J.$$

It is easy to see that $\ker \varphi = f^{-1}(J)$. By 1st isomorphism theorem, we obtain an injective homomorphism

$$\bar{\varphi} : R/f^{-1}(J) \hookrightarrow S/J.$$

Now, by Proposition 10.4.3, J being a prime implies that S/J is an integral domain. So $R/f^{-1}(J)$, realized as a subring of S/J, is automatically an integral domain. So $f^{-1}(J)$ is a prime ideal. □

10.6. **Principal ideal domains.** Initial study of rings is modeled on properties of $\mathbb{Z}$, trying to generalize various aspects of $\mathbb{Z}$ to other rings, such as certain quadratic rings, e.g. $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$.

Definition 10.6.1. A **principal ideal domain**, writing **PID** for short, is an integral domain in which every ideal is principal.

Example 10.6.2. (1) The ring of integers, $\mathbb{Z}$, is a PID. All of its ideals are of the form $n\mathbb{Z}$ for some n.

(2) For k a field, the ring of polynomials in one variable $k[x]$ is a PID.

(3) We will prove later that the ring $\mathbb{Z}[i]$ is a PID. This is a very important example.

(4) (Non-example) The ring $\mathbb{Z}[\sqrt{-5}]$ is not a principal ideal domain. For example, $(3, 1+2\sqrt{-5})$ is not a principal ideal.

Proposition 10.6.3. *Every nonzero prime ideal in a PID is a maximal ideal.*

Proof. Let (p) be a prime ideal in a PID R. If $\mathfrak{m} = (m) \supseteq (p)$ is a maximal ideal containing (p). Then $p = mn$ for some $n \in R$. Using the property of prime ideals, we see that either m or n belongs to (p).

- If $m \in (p)$, then $(m) \subseteq (p)$. This says that $(p) = (m)$.
- If $n \in (p)$, then $n = ps$ for some $s \in R$. Thus we have $p = mn = mps$, and thus $1 = ms$. This implies that m is a unit. So $\mathfrak{m} = (1)$ contradicting with that $\mathfrak{m}$ is a maximal ideal.

□

10.7. **Quadratic integer rings.**

10.7.1. *Quadratic integer rings.* Fix a square-free integer D (positive or negative), i.e. $D = \pm$ product of distinct primes, and $D \neq 1$. Let

$$\mathbb{Q}(\sqrt{D}) = \{x + y\sqrt{D} \,|\, x, y \in \mathbb{Q}\}$$

be the "quadratic field" (it is a two-dimensional $\mathbb{Q}$-vector space).

The "correct" analogue of $\mathbb{Z}$ inside $\mathbb{Q}(\sqrt{D})$ is

$$\mathcal{O} = \mathcal{O}_{\mathbb{Q}(\sqrt{D})} := \begin{cases} \mathbb{Z}[\sqrt{D}] & \text{if } D \equiv 2, 3 \bmod 4; \\ \mathbb{Z}[\frac{1+\sqrt{D}}{2}] & \text{if } D \equiv 1 \bmod 4. \end{cases}$$

This is because for $z = \sqrt{D}$ or $\frac{1+\sqrt{D}}{2}$, it is a zero of $z^2 - D$ or $z^2 - z + \frac{1-D}{4} \in \mathbb{Z}[z]$ in two cases.

One sometimes write this in a diagram as:

$$\begin{tikzcd}
\mathbb{Q}(\sqrt{D}) \arrow[d, dash] & \supset & \mathcal{O}_{\mathbb{Q}(\sqrt{D})} \arrow[d, dash] \\
\mathbb{Q} & \supset & \mathbb{Z}.
\end{tikzcd}$$

Notation 10.7.2. On $\mathbb{Q}(\sqrt{D})$, we define a **conjugation** map: $\overline{x + y\sqrt{D}} := x - y\sqrt{D}$. Clearly, it preserves the subring $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$, and satisfies

$$\overline{a \cdot b} = \bar{a} \cdot \bar{b}.$$

(In fact $a \mapsto \bar{a}$ is an automorphism of $\mathbb{Q}(\sqrt{D})$ and of $\mathcal{O}_{\mathbb{Q}(\sqrt{D})}$.) When $D < 0$, this is nothing but just the complex conjugation. But when $D > 0$, this makes sense purely because of algebraic structure we have on $\mathcal{O}$.

Define the following **norm map**:

$$\mathrm{Nm} : \mathbb{Q}(\sqrt{D}) \longrightarrow \mathbb{Q}$$
$$\mathrm{Nm}(x + y\sqrt{D}) := (x + y\sqrt{D})(x - y\sqrt{D}) = x^2 - Dy^2.$$

(If $D < 0$, $\mathrm{Nm}(a) \geq 0$ for any $a \in \mathbb{Q}(\sqrt{D})$.)

Properties 10.7.3. (1) If $x + y\sqrt{D} \in \mathcal{O}$, then $\mathrm{Nm}(x + y\sqrt{D}) \in \mathbb{Z}$.

(2) $\mathrm{Nm}(a) = a\bar{a}$ for any $a \in \mathbb{Q}(\sqrt{D})$.

(3) Nm is multiplicative, i.e. $\mathrm{Nm}(ab) = \mathrm{Nm}(a)\mathrm{Nm}(b)$ for $a, b \in \mathcal{O}_{\mathbb{Q}(\sqrt{D})}$. (Again, this is clear when $D < 0$ from the usual story of complex numbers, but when $D > 0$, this follows from the purely algebraic argument.

Proof. We leave (1) and (2) as exercises. For (3), we use (2) to note that $\mathrm{Nm}(ab) = ab\overline{ab} = a\bar{a} \cdot b\bar{b} = \mathrm{Nm}(a)\mathrm{Nm}(b)$. □

Lemma 10.7.4. *For an element $u \in \mathcal{O}$, $u \in \mathcal{O}^\times$ if and only if* $\mathrm{Nm}(u) = \pm 1$.

Proof. "⇐" As $\mathrm{Nm}(u) = u\bar{u} = \pm 1$, we have $u \in \mathcal{O}^\times$.

"⇒" If $uv = 1$ for some $v \in \mathcal{O}$, then

$$\mathrm{Nm}(u)\mathrm{Nm}(v) = \mathrm{Nm}(uv) = \mathrm{Nm}(1) = 1.$$

We must have $\mathrm{Nm}(u) \in \{\pm 1\}$. □

10.7.5. *Pell's equation.* When $D \equiv 2, 3 \bmod 4$, we have

$$x \pm y\sqrt{D} \in \mathcal{O}^\times \Leftrightarrow \mathrm{Nm}(x \pm y\sqrt{D}) = \pm 1 \Leftrightarrow x^2 - Dy^2 = \pm 1.$$

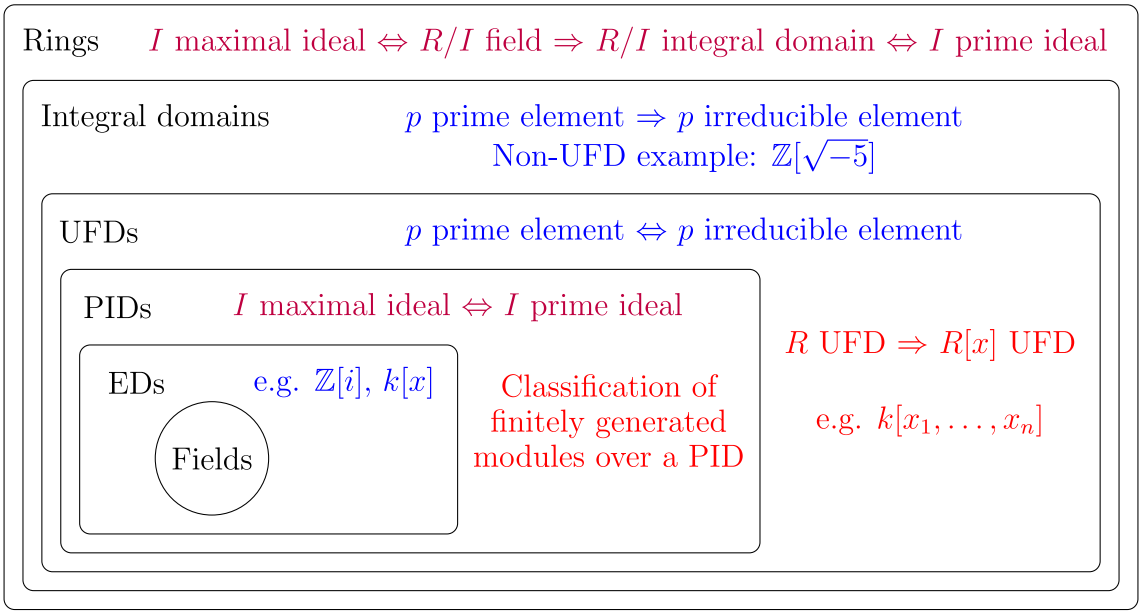
Thus, solutions of Pell's equation form the group $\mathcal{O}^\times$!

Fact 10.7.6. (1) When $D > 0$, $\mathcal{O}^\times = \left\{ \pm (x_0 + y_0\sqrt{D})^{\mathbb{Z}} \right\}$ for a "fundamental" unit $x_0 + y_0\sqrt{D} \in \mathcal{O}^\times$.

(2) When $D < 0$, we have $\mathcal{O}^\times = \{\pm 1\}$ unless
- when $D = -1$, $\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}$;
- when $D = -3$, $\mathbb{Z}[\zeta_3]^\times = \{\pm 1, \pm\zeta_3, \pm\zeta_3^2\}$ with $\zeta_3 = e^{2\pi i/3}$ is a nontrivial third root of unity.

11. Unique factorization domains

For this entire lecture, R will be an integral domain. Our goal is summarized into the following picture



Rings
I maximal ideal $\Leftrightarrow$ R/I field $\Rightarrow$ R/I integral domain $\Leftrightarrow$ I prime ideal
Integral domains
p prime element $\Rightarrow$ p irreducible element
Non-UFD example: $\mathbb{Z}[\sqrt{-5}]$
UFDs
p prime element $\Leftrightarrow$ p irreducible element
PIDs
I maximal ideal $\Leftrightarrow$ I prime ideal
EDs
e.g. $\mathbb{Z}[i]$, $k[x]$
Fields
Classification of finitely generated modules over a PID
R UFD $\Rightarrow$ $R[x]$ UFD
e.g. $k[x_1, \ldots, x_n]$

Here the purple colored statements were proved in previous lectures; we will prove the blue colored statements in this lecture, and the red colored statements in the following lectures.

11.1. Euclidean domains.

There is a question we hope to answer: how to prove that an integral domain is a PID? In classical theory of rings, one verifies this by checking a certain algorithm of finding generators of ideals on R.

Definition 11.1.1. An integral domain R is said to be an **Euclidean domain** if there is a norm $\mathrm{Nm} : R \to \mathbb{Z}^+ \cup \{0\}$ such that

(1) $\mathrm{Nm}(0) = 0$;
(2) for any $a, b \in R$ with $b \neq 0$, there exists "quotient" $q \in R$ and "remainder" $r \in R$ such that

$$a = bq + r \quad \text{and} \quad \text{either } r = 0 \text{ or } \mathrm{Nm}(r) < \mathrm{Nm}(b).$$

It is important to point out that *we do not require q and r to be unique.*

Remark 11.1.2. Recall that this Euclidean algorithm can be used to find the gcd of two integers. Euclidean domain is a ring in which such an algorithm is valid.

Example 11.1.3. (1) For any field F, we may take $N : F \to \mathbb{Z}_{\geq 0}$ by $N(a) = 0$.
(2) For a field F and $R = F[x]$ a polynomial ring, define $\mathrm{Nm}(f(x)) = \deg(f)$.
(3) The **ring of Gaussian integers** $R = \mathbb{Z}[i]$ admits a norm

$$\mathrm{Nm}(x + yi) = x^2 + y^2 = |x + yi|^2.$$

When $a, b \in \mathbb{Z}[i]$ with $b \neq 0$, let $q \in \mathbb{Z}[i]$ is taken so that $\big|\mathrm{Re}(q - \frac{a}{b})\big| \leq \frac{1}{2}$ and $\big|\mathrm{Im}(q - \frac{a}{b})\big| \leq \frac{1}{2}$, then

$$\mathrm{Nm}(a - bq) = |b|^2 \cdot \Big|\frac{a}{b} - q\Big|^2 \leq |b|^2 \cdot \Big(\frac{1}{4} + \frac{1}{4}\Big) < |b|^2.$$

(4) For $R = \mathbb{Z}[\zeta_3]$ with $\zeta_3 = e^{2\pi i/3}$, using the quadratic norm $\mathrm{Nm} : R \to \mathbb{Z}_{\geq 0}$ given by $\mathrm{Nm}(z) = z\bar{z}$, R is also an Euclidean domain.

Proposition 11.1.4. *A Euclidean domain R is a PID.*

Proof. Let $I \subseteq R$ be a nonzero ideal. Let b be an element of $I\backslash\{0\}$ with minimal possible norm. We claim that $I = (b)$. It is clear that $(b) \subseteq I$. Now we focus on the other inclusion.

If $a \in I$, by Euclidean algorithm, $a = bq + r$ with $r = 0$ or $\mathrm{Nm}(r) < \mathrm{Nm}(b)$. If $r \neq 0$, then $r = a - bq \in I$ contradicting that b has minimal norm. Thus $r = 0$ and thus $a \in (b)$. The theorem is proved. □

Remark 11.1.5. Unfortunately, beyond the examples above, there are only a few more examples of Euclidean domains. This concept is historically very important, but to establish more general PID properties, one needs more commutative algebra tools which we will learn in later courses.

11.2. **Generalization of prime number to a general integral domain.** As we explained earlier, our initial study of rings is to imitate what happens in $\mathbb{Z}$ (and maybe test that in the ring $\mathbb{Z}[i]$.

In $\mathbb{Z}$, prime numbers play an important role; we hope to generalize that here.

Definition 11.2.1. (1) For $a, b \in R$ with $a \neq 0$, we write $a \,|\, b$ if $b = ac$ for some $c \in R$. This is equivalent to say that $b \in (a)$.

(2) A nonzero element $p \in R$ is called a **prime element** if (p) is a prime ideal, or equivalently,

$$\text{if } p \,|\, ab, \text{ then } p \,|\, a \text{ or } p \,|\, b.$$

(3) Suppose that $r \in R$ is nonzero and not a unit. Then r is called an **irreducible element** if

$$\text{whenever } r = ab, \text{ then } a \text{ or } b \text{ is a unit.}$$

(4) Two elements $a, b \in R$ are said to be **associated** if $a = bu$ for some unit $u \in R^\times$. (This of course implies that $b = au^{-1}$ with $u^{-1} \in R^\times$ a unit.)

Proposition 11.2.2. (1) *Prime elements are always irreducible.*

(2) *If R is a PID, then irreducible elements are prime elements.*

Proof. (1) Let $p \in R$ be a prime element. Then if $p = uv$ for some $u, v \in R$, then

$$uv \in (p) \;\Rightarrow\; u \in (p) \text{ or } v \in (p).$$

WLOG, $u = ps$ for some $s \in R$, then $p = uv = psv$. Thus $1 = sv$ and thus v is a unit. So p is irreducible.

(2) If p is irreducible, we hope to show that (p) is a prime ideal; in fact we will show that (p) is a maximal ideal.

Indeed, if $(p) \subseteq (m)$ for another ideal with $m \in R$, then $p = r \cdot m$ for some $r \in R$. By the property of irreducible elements,

- either r is a unit, which implies that $(p) = (m)$;
- or m is a unit, which implies that $(m) = (1)$.

So (p) is a maximal ideal and hence a prime ideal. So p is a prime element.

□

Definition 11.2.3. A **unique factorization domain** (**UFD** for short) is an integral domain R in which every nonzero and non-unit element $r \in R$ satisfies that

(1) r is a product of irreducibles, namely $r = p_1 p_2 \dots p_n$ with $p_i \in R$ irreducible elements; and
(2) the decomposition in (1) is **unique up to associates**, namely, if $r = q_1 \cdots q_m$ is another factorization of r into product of irreducible elements, then $m = n$ and there exists $\sigma \in S_n$ such that p_n and $q_{\sigma(n)}$ are associates.

Example 11.2.4. The rings $\mathbb{Z}$ and $k[x_1, \dots, x_n]$ for a field k are UFDs.

A typical non-UFD integral domain is $R = \mathbb{Z}[\sqrt{-5}] = \{x + y\sqrt{-5} \,|\, x, y \in \mathbb{Z}\}$. We will give more discussion in § 11.5.

Remark 11.2.5. Why do we need to require factors to be unique *up to associates*? This already happens in $\mathbb{Z}$, for example, $6 = 2 \cdot 3 = (-2) \cdot (-3)$. They are essentially the same factorization if one redistribute the unit factor -1 among the irreducible factors.

Two main theorem we will prove later in this lecture is:

(1) PID $\Rightarrow$ UFD, and
(2) If R is a UFD, so is $R[x_1, \dots, x_n]$.

Proposition 11.2.6. *In a UFD R, for a nonzero element $p \in R$, p is a prime element $\Leftrightarrow$ p is an irreducible element.*

Proof. "$\Rightarrow$" This is true in any integral domain by Proposition 11.2.2(1).

"$\Leftarrow$" If $p \,|\, ab$, then $ab = pc$ for some $c \in R$. Writing a and b as product of irreducible elements, we must find an associate of p in the product of such expression by the uniqueness of factorization. Thus $a = pr$ or $b = pr$ for some $r \in R$. This means that $p \,|\, a$ or $p \,|\, b$. □

Proposition 11.2.7. *In a UFD, the gcd of nonzero element exists. Namely, for two nonzero element $a, b \in R$, there exists an element $d = \gcd(a, b) \in R$ such that if $d' \in R$ satisfies $d' \,|\, a$ and $d' \,|\, b$, then $d' \,|\, d$.*

Explicitly, if a and b factor as

$$a = up_1^{c_1} \cdots p_r^{c_r} \quad \text{and} \quad b = vp_1^{d_1} \cdots p_r^{d_r}$$

with $p_1, \dots, p_r$ irreducible and pairwise non-associate, and $u, v \in R^\times$, $c_i, d_i \in \mathbb{Z}_{\geq 0}$, then

$$d = p_1^{\min(c_1,d_1)} \cdots p_r^{\min(c_r,d_r)}$$

is a gcd of a and b.

Remark 11.2.8. The gcd of two elements of R is unique up to associates.

11.3. **PID $\Rightarrow$ UFD.**

Theorem 11.3.1. *If R is a PID, then R is a UFD.*

Proof. Existence of factorization. Let $r \in R$ be a nonzero element that is not a unit.

- If r is irreducible, then we are done.
- Otherwise, $r = a_1 \cdot b_1$ for a_1 and b_1 non-unit.

We may continue this process for a_1 and b_1, respectively.

Suppose that this process does not terminate. Then we may keep writing

$$r = a_1 b_1 = a_1 a_2 b_2 = a_1 a_2 a_3 b_3 = \cdots$$

Thus, we have

$$(r) \subseteq (b_1) \subseteq (b_2) \subseteq \cdots$$

We need to use axiom of choice in this step.[4] Taking the union

$$\bigcup_{n \geq 0} (b_n) = \text{ some ideal } (b).$$

But this element b must be contained in one of (b_n) and thus $(b_n) = (b_{n+1}) = \cdots$.

In particular, $b_n = a_{n+1} b_{n+1}$ and $(b_n) = (b_{n+1})$ implies that a_{n+1} is a unit. This is a contradiction.

Uniqueness of the decomposition. We make induction on the number of irreducible factors. If $n = 0$, then r is a unit. If $r = qc$ for some irreducible element q, then $q \,|\, 1$ and hence q is a unit, giving a contradiction. So in the factorization of r, there cannot be any irreducible element.

Now suppose that $n \geq 1$ and $r = p_1 p_2 \cdots p_n = q_1 \cdots q_m$ with $m \geq n$ and p_i and q_j irreducible elements. Then p_1 divides $q_1 \cdots q_m$ and hence divides one of them (because by Proposition 11.2.2(2), p_1 is a prime element).

WLOG, we have $p_1 \,|\, q_1$. This implies that $q_1 = p_1 u$ for u a unit (because q_1 is irreducible). This implies that

$$(u^{-1} p_2) \cdot p_3 \cdots p_n = q_2 \cdots q_m.$$

By inductive hypothesis, we are done. □

We emphasize again the implications: Euclidean domain $\Rightarrow$ PID $\Rightarrow$ UFD.

Lemma 11.3.2. *Let R be a PID and a, b be nonzero elements in R. Then $(a, b) = (\gcd(a, b))$.*

Proof. Since R is a PID, the ideal (a, b) is principal, so equals to (d) for some $d \in R$. This in particular says that $a, b \in (d)$, i.e. $d|a$ and $d|b$. So $d| \gcd(a, b)$. On the other hand, by definition, $d = xa + yb$ for some $x, y \in R$. Yet $\gcd(a, b)$ divides a and b, so $\gcd(a, b)$ divides d. So d and $\gcd(a, b)$ are associates. The lemma is proved. □

[4] At least, we need an axiom of countably infinite dependent choice. A homogeneous relation R on X is called a *total relation* if for every $a \in X$, there exists some $b \in X$ such that $a \, R \, b$ is true. The axiom of (countably infinite) dependent choice is: for every nonempty set X and every total relation R on X, there exists a sequence $(x_n)_{n \in \mathbb{N}}$ in X such that $x_n \, R \, x_{n+1}$ for all $n \in \mathbb{N}$.

11.4. **Application to Gaussian integers.** Take $R = \mathbb{Z}[i]$. It is an Euclidean domain and hence a PID and a UFD. Consider the norm map

$$\begin{gathered}\mathrm{Nm} : \mathbb{Z}[i] \longrightarrow \mathbb{Z}_{\geq 0},\\ \mathrm{Nm}(x+yi) = x^2+y^2 = |x+iy|^2.\end{gathered}$$

The group of units in $\mathbb{Z}[i]$ is $\mathbb{Z}[i]^\times = \{a \in \mathbb{Z}[i] \,|\, \mathrm{Nm}(a) = 1\} = \{\pm 1, \pm i\}$.

Theorem 11.4.1. (1) *(Fermat's theorem on sums of squares) A prime p is the sum of two square of integers (i.e. $p = x^2 + y^2$ for $x, y \in \mathbb{Z}$) if and only if $p = 2$ or $p \equiv 1 \bmod 4$.*

Moreover, such x and y's are unique up to swapping x with y and changing signs.

(2) *Irreducible elements in $\mathbb{Z}[i]$ are as follows (up to associates):*
   (a) $1+i$ *(with norm 2);*
   (b) *the primes $p \in \mathbb{Z}$ such that $p \equiv 3 \bmod 4$ (with norm p^2); and*
   (c) *$x + yi$ and $x - yi$, if $p = x^2 + y^2$ for $x, y \in \mathbb{Z}$, for a prime $p \equiv 1 \bmod 4$, (with norm p).*

Proof. Step 1: If $\pi \in \mathbb{Z}[i]$ is so that $\mathrm{Nm}(\pi)$ is a prime number p, then π is irreducible.

Indeed, if $\pi = ab$, then $p = \mathrm{Nm}(\pi) = \mathrm{Nm}(a)\mathrm{Nm}(b)$. So either $\mathrm{Nm}(a) = 1$ or $\mathrm{Nm}(b) = 1$; by Lemma 10.7.4, we have either a or b is a unit.

Step 2: For every irreducible element $\pi \in \mathbb{Z}[i]$, $\mathrm{Nm}(\pi) = p$ or p^2 for some prime p.

We look at the intersection $(\pi) \cap \mathbb{Z}$. It is a prime ideal in $\mathbb{Z}$ because if $a, b \in \mathbb{Z}$ satisfies $ab \in (\pi) \in \mathbb{Z}$, then either $a \in (\pi)$ or $b \in (\pi)$ as (π) is a prime ideal. Thus $a \in (\pi) \cap \mathbb{Z}$ or $b \in (\pi) \cap \mathbb{Z}$. So $(\pi) \cap \mathbb{Z}$ is a prime ideal and thus $(\pi) \cap \mathbb{Z} = (p)$ for some prime p.

This implies that $p = \pi a$ for some $a \in \mathbb{Z}[i]$. Yet

$$p^2 = \mathrm{Nm}(p) = \mathrm{Nm}(\pi)\mathrm{Nm}(a).$$

We separate two possibilities.

- If $\mathrm{Nm}(\pi) = p^2$, then $\mathrm{Nm}(a) = 1$ and thus $a \in \{\pm 1, \pm i\}$. So π is associated to p.
- If $\mathrm{Nm}(\pi) = p$, then $p = \pi\bar{\pi}$ and both π and $\bar{\pi}$ are irreducible elements in $\mathbf{Z}[i]$.

Step 3: We study the two possibilities above in terms of $p \bmod 4$.

- If $p = 2$, $2 = (1+i)(1-i)$. Yet $1 - i = -i(1+i)$ is associated to $1+i$.
- If $p \equiv 3 \bmod 4$, p is irreducible in $\mathbb{Z}[i]$, otherwise, $p = \mathrm{Nm}(\pi) = a^2 + b^2$ for some $a, b \in \mathbb{Z}$. But $a^2 + b^2 \equiv 0, 1, 2 \bmod 4$. This is a contradiction.
- If $p \equiv 1 \bmod 4$, we *hope* to show that $p = \pi\bar{\pi}$ for some $\pi = a + bi$ irreducible, and thus $p = (a+bi)(a-bi) = a^2 + b^2$, proving the theorem. Thus, it suffices to show that p is not irreducible in $\mathbb{Z}[i]$.

  We make use of a fact that $\mathbf{Z}_p^\times$ is a cyclic group of order $p - 1$ (which is a multiple of 4). This fact will be proved later in this semester. Admitting this fact, we see that there exists $a \in \mathbf{Z}_p^\times$ such that

  $$a^4 \equiv 1 \bmod p \quad \text{yet} \quad a^2 \not\equiv 1 \bmod p.$$

  This implies that $a^2 + 1 \equiv 0 \bmod p$.

  If p was irreducible in $\mathbf{Z}[i]$, then $p \,|\, a^2 + 1 = (a+i)(a-i)$. Thus either $p \,|\, a + i$ or $p \,|\, a - i$. But clearly the "coefficients on i is 1 and is not divisible by p. This is a contradiction, and thus p is not irreducible in $\mathbb{Z}[i]$.

□

Extended readings after Section 13

11.5. **Beyond PIDs and UFDs.** We shortly investigate what happens if the unique factorization property fails, and hopefully points to how this issue was handled later, mostly in number theory.

Let us take an example $R = \mathbb{Z}[\sqrt{-5}]$, equipped with a norm map

$$\mathrm{Nm} : R \to \mathbb{Z}_{\geq 0}, \qquad \mathrm{Nm}(x + y\sqrt{-5}) = x^2 + 5y^2.$$

In particular, by judging from the formula of the norm, we see that there is no element in R with norm 2, 3, 7, and etc.

Now, consider

$$21 = 3 \times 7 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5}). \tag{11.5.0.1}$$

We claim that each of 3, 7, and $1 \pm 2\sqrt{-5}$ is an irreducible element.

Indeed, $\mathrm{Nm}(3) = 9$, $\mathrm{Nm}(7) = 49$ and $\mathrm{Nm}(1 \pm 2\sqrt{-5}) = 21$. But we always have $\mathrm{Nm}(ab) = \mathrm{Nm}(a)\mathrm{Nm}(b)$, yet no elements of R have norm 3 or 7.

Thus, R is not a UFD and hence not a PID.

11.6. **Replacement of UFD properties.** In number theory, instead of requiring unique factorization into irreducible elements, we require unique factorization into nonzero prime ideals.

An integral domain R is called a **Dedekind domain** if every nonzero ideal a unique factorization into a product of prime ideals.

A major class of examples of Dedekind domains are "ring of integers" $\mathcal{O}_K$, namely the analogue of $\mathbb{Z} \subset \mathbb{Q}$ for a finite extension of $\mathbb{Q}$.

$$\begin{tikzcd}
K = \mathbb{Q}(\alpha) \arrow[d, dash, "\text{finite extension}"'] & \mathcal{O}_K \arrow[l, hook'] \arrow[d, dash] \\
\mathbb{Q} & \mathbb{Z} \arrow[l, hook']
\end{tikzcd}$$

Here is one of the most famous example of such Dedekind domain. Consider an odd prime p, and set $\zeta_p = e^{2\pi i/p}$. Then

$$\begin{tikzcd}
K = \mathbb{Q}(\zeta_p) \arrow[d, dash] & \supset & \mathcal{O}_K \arrow[d, dash] & = & \mathbb{Z}[\zeta_p] = \left\{a_0 + a_1\zeta_p + \cdots + a_{p-2}\zeta_p^{p-2} \,\middle|\, a_i \in \mathbb{Z}\right\} \\
\mathbb{Q} & \supset & \mathbb{Z}.
\end{tikzcd}$$

Here is a "failed" approach to Fermat's Last Theorem: for a prime $p \geq 3$, the equation

$$x^p + y^p = z^p$$

has no nontrivial solutions. Suppose that there is a solution with $\gcd(x, y, z) = 1$, then

$$x^p = z^p - y^p = (z - y)(z - \zeta_p y) \cdots (z - \zeta_p^{p-1} y).$$

If $\mathbb{Z}[\zeta_p]$ is a PID (which is true when $p = 3$), then it is a UFD. We then essentially proved that

$$z - y,\ z - \zeta_p y,\ \cdots \text{are all relatively prime except possible at prime factors at } p.$$

So each of above is almost a pth power. One can imagine that this will provide many new information and a solution by infinite descent to Fermat's Last Theorem is within reach.

Unfortunately, the ring $\mathbb{Z}[\zeta_p]$ is not a UFD in general (and probably not a UFD for all primes $p \geq 5$). The good news is that $\mathbb{Z}[\zeta_p]$ is a Dedekind domain, which means that we may still deduce that the ideal generated by each of $z - \zeta_p^i y$ is almost a pth power of an ideal. In algebraic number theory, there is a concept called *ideal class group* $\mathrm{Cl}(\mathcal{O}_K)$ to measure how far $\mathcal{O}_K$ is from being a PID. It is a finite abelian group. One can prove that when $p \nmid |\mathrm{Cl}(\mathbb{Z}[\zeta_p])$, the Fermat's Last Theorem holds for p. Primes satisfying this condition are called regular primes. The first several irregular primes are $p = 37, 59, 67, 101, \ldots$. Although this did not solve Fermat's Last Theorem for all p, it has made a significant step forward.

12.1. **Polynomial rings over a UFD.** In this section, we prove the following main result.

Theorem 12.1.1. *An integral domain R is a UFD if and only if $R[x]$ is a UFD.*

Corollary 12.1.2. *If R is a UFD, then $R[x_1, \dots, x_n]$ is a UFD.*

Example 12.1.3. For example, starting with $R = \mathbb{Z}$ or $\mathbb{Z}[i]$ or a field k being a UFD, we deduce that $\mathbb{Z}[x_1, \dots, x_n]$, $\mathbb{Z}[i][x_1, \dots, x_n]$, and $k[x_1, \dots, x_n]$ are UFDs.

12.1.4. *Proof of Theorem 12.1.1.* Write $F = \mathrm{Frac}(R)$. Then $F[x]$ is an ED $\Rightarrow$ PID $\Rightarrow$ UFD.

Step 0: It is clear that a constant $a \in R$

- is a unit in R if and only if a is a unit in $R[x]$;
- is irreducible in R if and only if a is irreducible in $R[x]$. (This is because if $a = bc$ then b and c are constant polynomials.)

From this, we see that $R[x]$ is a UFD $\Rightarrow$ R is a UFD. We will prove that R UFD $\Rightarrow$ $R[x]$ UFD. In below, we will assume that R is a UFD, and will show that $R[x]$ is a UFD.

The constant functions are already treated. We now consider polynomials of degree ≥ 1. The key is to relate this with the situation in $F[x]$.

Step 1: (Gauss' Lemma) Let $p(x) \in R[x]$ be a nonzero polynomial. If $p(x)$ is reducible in $F[x]$, then $p(x)$ is reducible in $R[x]$, i.e. if $p(x) = A(x)B(x)$ for nonconstant polynomials $A(x), B(x) \in F[x]$, then there exists $r \in F^\times$ such that $a(x) = rA(x)$ and $b(x) = r^{-1}B(x)$ are both in $R[x]$.

Proof of Gauss' lemma: Let d be a l.c.m. of the denominators of coefficients of $A(x)$ and $B(x)$. This implies that

$$d \cdot p(x) = a_1(x)b_1(x) \quad \text{for} \quad a_1(x), b_1(x) \in R[x].$$

- If d is a unit in R, then $p(x) = (d^{-1}a_1(x)) \cdot b_1(x)$. We are done.
- Otherwise, take a prime factor q of d. Then $R[x]/(q) = (R/qR)[x]$ is an integral domain. (Using overline to denote the reduction modulo q,) note that

  $$0 = \overline{d \cdot p(x)} = \overline{a_1(x)} \cdot \overline{b_1(x)}.$$

  WLOG, assume that $\overline{a_1(x)} = 0$. This implies that all coefficients of $a_1(x)$ are divisible by q. So write

  $$d = qd_2, \quad a_2(x) = q^{-1}a_1(x) \in R[x], \quad \text{and} \quad b_2(x) = b_1(x).$$

  This says that

  $$d_2 p(x) = a_2(x)b_2(x)$$

  and we may continue with the discussion.

Step 2: Prove that the irreducible elements in $R[x]$ consists of

- constants $a \in R$ such that a is irreducible in R;
- polynomials $a(x) \in R[x]$ of degree ≥ 1, such that
  (i) $\gcd\big(\text{coefficients of } a(x)\big) = 1$, and
  (ii) $a(x)$ is irreducible in $F[x]$.

Proof: The case of constants follow from Step 0. Now, for $a(x) \in R[x]$ of degree ≥ 1,

- if $\gcd\big(\text{coefficients of } a(x)\big) = g$ is not a unit, then $a(x) = g \cdot (g^{-1}a(x))$ is not irreducible;
- if $a(x) = A(x)B(x)$ is reducible in $F[x]$, then Step 1 implies that $a(x)$ is reducible in $R[x]$.

Conversely, if $a(x)$ satisfies (i) and (ii) and $a(x) = b(x)c(x)$ in $R[x]$, then

- if $\deg b(x) \geq 1$ and $\deg c(x) \geq 1$, then (ii) does not hold, and
- if otherwise, WLOG $\deg b(x) = 0$. But if $b \in R$ is not a unit, then (i) does not hold. Thus b is a unit in R. We are done.

Step 3: Existence of the factorization in $R[x]$.

Given $a(x) \in R[x]$, if $a(x)$ is a constant, then we are reduced to the UFD property of R.

Now $\deg a(x) \geq 1$. Put $d := \gcd\big(\text{coefficients of } a(x)\big)$. Thus,

$$a(x) = da_1(x) \quad \text{for some } a_1(x) \in R[x],$$

where d may be factored into products of irreducibles.

On the other hand, we may factor $a_1(x) = A_1(x) \cdots A_r(x)$ in $F[x]$. By Gauss' lemma, we may adjust these factors so that each $A_i(x) \in R[x]$. Moreover, for each i, $\gcd\big(\text{coefficients of } A_i(x)\big) = 1$, otherwise $\gcd\big(\text{coefficients of } a_1(x)\big) \neq 1$, contradiction!

Thus, all $A_i(x)$ are irreducible elements.

Step 4 Uniqueness of the factorization.

Suppose that $a(x) = p_1(x) \cdots p_r(x) = q_1(x) \cdots q_s(x)$ are two factorizations into irreducibles.

We first view this factorization in $F[x]$. Each $p_i(x)$ with degree ≥ 1 must be associated to a $q_j(x)$ in $F[x]$, namely

$$p_i(x) = r \cdot q_j(x) \quad \text{for some} \quad r = \frac{a}{b} \in F^\times \text{ with } \gcd(a, b) = 1.$$

This implies that

$$bp_i(x) = aq_j(x).$$

- $\gcd\big(\text{coefficients of } q_j(x)\big)$ is divisible by b. Thus b must be a unit;
- $\gcd\big(\text{coefficients of } p_i(x)\big)$ is divisible by a. Thus a must be a unit.

From this discussion, we deduce that $p_i(x) = c \cdot q_i(x)$ for some unit $c \in R$.

Removing this pair of factors from the factorizations of $a(x)$, we are reduced to two factorizations of $\frac{a(x)}{p_i(x)}$. Repeating this process, we are reduced to the case when $a(x) \in R \subseteq R[x]$. Now, the uniqueness of factorization follows from that of R. $\square$.

12.2. **Irreducible criterion for polynomials.** In this subsection, we discuss criteria to test whether a polynomial is irreducible or not.

Lemma 12.2.1. (1) *If F is a field, a polynomial $f \in F[x]$ of degree 2 or 3 is irreducible if and only if it has a root in F.*

(2) *If $p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{Z}[x]$ is a polynomial such that $p(\frac{r}{s}) = 0$ for $r, s \in \mathbb{Z}$ with $\gcd(r, s) = 1$, then $r|a_0$ and $s|a_n$.*

Proof. (1) is clear. We prove (2). Suppose that

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

$$a_nr^n + a_{n-1}r^{n-1}s + \cdots a_0s^n = 0$$

This implies that

$$r \,\big|\, a_0s^n \quad \text{and} \quad s \,\big\|\, a_nr^n.$$

Thus, we deduce that $r|a_0$ and $s|a_n$. □

Example 12.2.2. The polynomial $f(x) = x^3 - x - 2 \in \mathbb{Z}[x]$ is irreducible because ± 1 and ± 2 are not zeros of $f(x)$.

Proposition 12.2.3 (Eisenstein's criterion)**.** *Let $\mathfrak{p}$ be a prime ideal of an integral domain R, and let $f(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_0 \in R[x]$. Suppose that*

(1) $c_0, c_1, \ldots, c_{n-1} \in \mathfrak{p}$, *and*
(2) $c_0 \notin \mathfrak{p}^2$.

Then $f(x)$ is irreducible.

Proof. We may assume that $\deg f(x) \geq 2$. Suppose that $f(x) = a(x)b(x)$ with $\deg a \geq 1$ and $\deg b \geq 1$. Then the leading coefficients of $a(x)$ and $b(x)$ are both units. So we may rescale $a(x)$ and $b(x)$ so that both polynomials are monic.

Taking the equation $f(x) = a(x)b(x)$ modulo $\mathfrak{p}$, we get

$$x^n = \bar{f}(x) = \bar{a}(x) \cdot \bar{b}(x)$$

in $R/\mathfrak{p}[x]$.

We claim that the constant terms $\bar{a}_0$ and $\bar{b}_0$ of $a(x)$ and $b(x)$ are zero. Indeed, if $\bar{a}_0 \neq 0$ and i is the minimal number such that $\bar{b}_i \neq 0$. Then in the product $\bar{a}(x)\bar{b}(x)$ we have a term $\bar{a}_0\bar{b}_ix^i$, contradicting with $x^n = \bar{a}(x)\bar{b}(x)$.

Thus $\bar{a}_0 = \bar{b}_0 = 0$. It follows that $a_0, b_0 \in \mathfrak{p}$. Thus the constant coefficient $c_0 \in \mathfrak{p}^2$, contradicting to (2). □

Example 12.2.4. The typical application of Eisenstein's criterion is to the cyclotomic polynomial. Let p be a prime number, the **pth cyclotomic polynomial** is

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

We claim that $\Phi_p(x)$ is irreducible in $\mathbb{Z}[x]$. This is because we consider $\Phi_p(x+1)$ instead:

$$\Phi_p(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = x^p + \underbrace{px^{p-1} + \binom{p}{2}x^{p-2} + \cdots + p}_{\text{all terms divisible by } p}$$

By Eisenstein's criterion, $\Phi_p(x+1)$ is irreducible, and hence $\Phi_p(x)$ is irreducible.

We remark that the above argument is related to the ideal identity

$$p\mathbb{Z}[\zeta_p] = (\zeta_p - 1)^{p-1} \quad \text{in} \quad \mathbb{Z}[\zeta_p].$$

We say that "the prime p is totally ramified in $\mathbb{Z}[\zeta_p]$."

12.3. **Factorization of polynomial ring quotient by a polynomial.**

Lemma 12.3.1. *Let F be a field. A polynomial $f(x)$ is irreducible if and only if $F[x]/(f(x))$ is a field.*

Proof. This follows from the following equivalences.

$$\begin{aligned} f(x) \text{ is an irreducible polynomial} \quad & \overset{F[x] \text{ UFD}}{\iff} \quad f(x) \text{ is a prime element in } F[x] \\ & \iff \quad (f(x)) \text{ is a prime ideal in } F[x] \\ & \overset{F[x] \text{ is a PID}}{\iff} \quad (f(x)) \text{ is a maximal ideal in } F[x] \\ & \iff \quad F[x]/(f(x)) \text{ is a field.} \end{aligned}$$

□

Example 12.3.2. This lemma will be useful later when we construct field extensions.

For example, the polynomial $x^3 + 2x - 1$ is irreducible in $\mathbb{F}_3[x]$ (think about why?) then

$$\mathbb{F}_3[x]/(x^3 + 2x - 1) \cong \big\{a + bx + cx^2 \,\big|\, a, b, c \in \mathbb{F}_3\big\}$$

is a field of 27 elements. (Will prove that there is a unique such field, up to isomorphism.)

Lemma 12.3.3. *Let F be a field. If $f(x) = p_1(x)^{n_1} \cdots p_r(x)^{n_r}$ is the factorization of $f(x)$ in $F[x]$, then*

$$\frac{F[x]}{(f(x))} \cong \frac{F[x]}{(p_1(x)^{n_1})} \times \cdots \times \frac{F[x]}{(p_r(x)^{n_r})}.$$

Proof. As the $p_i(x)^{n_i}$'s are pairwise coprime, i.e. $(p_i(x)^{n_i}, p_j(x)^{n_j}) = \big(\gcd(p_i(x)^{n_i}, p_j(x)^{n_j})\big) = (1)$, this follows from Chinese remainder theorem. □

Lemma 12.3.4. *Let F be a field. If $f(x) \in F[x]$ has distinct zeros $\alpha_1, \dots, \alpha_n$, then $f(x)$ is divisible by $(x - \alpha_1) \cdots (x - \alpha_n)$ in $F[x]$.*

In particular, a degree n polynomial can have at most n zeros.

Corollary 12.3.5. *If F is a field and G a finite subgroup of F, then G is cyclic.*

In particular, if F is a finite group, then $F^\times$ is cyclic.

Proof. Assume that $|G| = n$. By classification of finite abelian groups, we may write

$$G = \mathbf{Z}_{n_1} \times \mathbf{Z}_{n_2} \times \cdots \times \mathbf{Z}_{n_r} \quad \text{with integers} \quad n_1 \,|\, n_2 \,|\, \cdots \,|\, n_r \text{ and } n = n_1 \cdots n_r.$$

If G is not cyclic, all elements would have order dividing $n_r < n$, i.e. for any $g \in G$, $g^{n_r} = 1$. But this would mean that the polynomial $x^{n_r} - 1$ has n zeros in F. This is a contradiction! □

12.3.6. *Structure of* $(\mathbb{Z}/n\mathbb{Z})^\times$. We determine the structure of $(\mathbb{Z}/n\mathbb{Z})^\times$ as follows.

Step 1: Assume that n factors as $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$. The Chinese remainder theorem implies an isomorphism

$$\mathbb{Z}/n\mathbb{Z} \cong (\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_r^{\alpha_r}\mathbb{Z})$$

as rings. Taking the units in these two rings, we obtain an isomorphism of abelian groups.

$$\big(\mathbb{Z}/n\mathbb{Z}\big)^\times \cong \big(\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z}\big)^\times \times \cdots \times \big(\mathbb{Z}/p_r^{\alpha_r}\mathbb{Z}\big)^\times$$

(Note that a simple counting gives the usual formula for Euler's function $\phi(n) = \phi(n_1)\cdots\phi(n_r)$ when $n_1, \dots, n_r$ are pairwise coprime.)

Step 2 Fix a prime p, we show that

(1) for p an odd prime, $(\mathbb{Z}/p^\alpha\mathbb{Z})^\times$ is a cyclic group of order $p^{\alpha-1}(p-1)$;
(2) for $p = 2$, $(\mathbb{Z}/2\mathbb{Z})^\times = \{1\}$ and when $\alpha \geq 2$, $(\mathbb{Z}/2^\alpha\mathbb{Z})^\times \cong \mathbf{Z}_2 \times \mathbf{Z}_{2^{\alpha-2}}$.

To show this, we first note that there is a surjective homomorphism

(12.3.6.1)
$$\begin{tikzcd} \phi : (\mathbb{Z}/p^\alpha\mathbb{Z})^\times \arrow[r, "\mathrm{mod}\ p"] & (\mathbb{Z}/p\mathbb{Z})^\times \\ a \arrow[r, mapsto] & a \bmod p. \end{tikzcd}$$

The target $(\mathbb{Z}/p\mathbb{Z})^\times$ is a cyclic group of order $p-1$. The kernel $\ker\phi = \big\{a \in \mathbb{Z}/p^\alpha\mathbb{Z} \,\big|\, a \equiv 1 \bmod p\big\}$ has order $p^{\alpha-1}$.

We claim that $\ker\phi$ is a cyclic group of order $p^{\alpha-1}$. Consider the following "exponential map" and the "logarithmic map"

(12.3.6.2)
$$\begin{tikzcd} \ker\phi = \big(1 + p\mathbb{Z}/p^\alpha\mathbb{Z},\ \cdot\big) \arrow[r, shift left, "\frac{1}{p}\log(-)"] & \mathbf{Z}_{p^{\alpha-1}} \arrow[l, shift left, "\exp(p-)"] \\ 1 + px \arrow[r] & \frac{1}{p}\log(1+px) = \frac{1}{p}\big(px - \frac{(px)^2}{2} + \frac{(px)^3}{3} - \cdots\big) \\ \exp(py) = 1 + py + \frac{p^2y^2}{2!} + \cdots & y \arrow[l, mapsto] \end{tikzcd}$$

One can easily check that these two maps are homomorphisms and are inverses of each other, giving rise to an isomorphism between $\ker\phi$ and $\mathbf{Z}_{p^{\alpha-1}}$.

Now, by the classification of finitely generated abelian groups, we may write

$$(\mathbb{Z}/p^\alpha\mathbb{Z})^\times = \ker\varphi \times G^p$$

for some group G^p that has order prime-to-p. Since φ of (12.3.6.1) maps $\ker\psi$ to $1 \in (\mathbb{Z}/p\mathbb{Z})^\times$, φ must induces an isomorphism between G^p with $(\mathbb{Z}/p\mathbb{Z})^\times \simeq \mathbf{Z}_{p-1}$. From this, we see that

$$(\mathbb{Z}/p^\alpha\mathbb{Z})^\times \cong \ker\phi \times G^p \simeq \mathbf{Z}_{p^{\alpha-1}} \times \mathbf{Z}_{p-1} \simeq \mathbf{Z}_{(p-1)p^{\alpha-1}}.$$

The case when $p = 2$ can be treated similarly, except that (12.3.6.2) should be replaced by

$$\begin{tikzcd} \ker\phi = \big(1 + 4\mathbb{Z}/2^\alpha\mathbb{Z},\ \cdot\big) \arrow[r, shift left, "\frac{1}{4}\log(-)"] & \mathbf{Z}_{2^{\alpha-2}} \arrow[l, shift left, "\exp(4-)"] \\ 1 + 4x \arrow[r] & \frac{1}{4}\log(1+4x) = \frac{1}{4}\big(4x - \frac{(4x)^2}{2} + \frac{(4x)^3}{3} - \cdots\big) \\ \exp(4y) = 1 + 4y + \frac{4^2y^2}{2!} + \cdots & y \arrow[l, mapsto] \end{tikzcd}$$

Here we need 4 instead of 2 so that $\exp(-)$ "converges". We may similarly deduce that, when $\alpha \geq 2$,

$$(\mathbb{Z}/4^\alpha\mathbb{Z})^\times \cong \{\pm 1\} \times \big(1 + 4\mathbb{Z}/2^\alpha\mathbb{Z},\ \cdot\big) \cong \{\pm 1\} \times \mathbf{Z}_{2^{\alpha-2}}.$$

We leave the details of the argument to the readers to fill in.

13. Modules and classification of finitely generated modules over PID

13.1. **Definition of modules.** Modules are analogues of vector spaces. We have learned that a vector space is an abelian group on which a field acts. A module is an abelian group on which a ring acts.

Definition 13.1.1. Let R be a ring (with $1_R \neq 0_R$). A **left R-omdule** is an abelian group M equipped with an R-action on M:

$$\begin{aligned} R \times M &\longrightarrow M \\ (a, m) &\longmapsto a \cdot m \end{aligned}$$

satisfying the following conditions (where $m, n \in M$ and $r, s \in R$)

(0) $1_R \cdot m = m$,
(1) $(r + s) \cdot m = r \cdot m + s \cdot m$,
(2) $r \cdot (m + n) = r \cdot m + r \cdot n$,
(3) $r \cdot (s \cdot m) = (rs) \cdot m$.

An **R-submodule** $N \subseteq M$ is an abelian subgroup N of M such that

$$\forall r \in R, \quad r \cdot N \subseteq N.$$

A **right R-module** is an abelian group M with a right R-action

$$\begin{aligned} M \times R &\longrightarrow M \\ (m, a) &\longmapsto m \cdot a \end{aligned}$$

satisfying analogues of (0), (1), and (2) above and

$(3)_R$ $(m \cdot s) \cdot r = m \cdot (sr)$.

Remark 13.1.2. If R is commutative, then there is no difference between left and right R-modules.

Example 13.1.3. (1) If $R = F$ is a field, then F-modules are the same as F-vector spaces.

(2) The free modules $R^{\oplus n} = \{(a_1, \dots, a_n) \,|\, a_i \in R\}$ with operations

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n) \quad \text{and} \quad r \cdot (a_1, \dots, a_n) = (ra_1, \dots, ra_n).$$

(3) If $I \subseteq R$ is a left ideal, then I is a left R-submodule of R. This is because for any $r \in R$, $r \cdot I \subseteq I$.

Conversely, a left R-submodule of R is a left ideal.

(4) A $\mathbb{Z}$-module is just an abelian group G. The operation is given by

$$\forall n \in \mathbb{N}, \quad n \cdot g = \underbrace{g + \cdots + g}_{n \text{ times}} \in G \quad \text{and} \quad (-n) \cdot g = -n \cdot g.$$

Similarly, a $\mathbb{Z}$-submodule is an abelian subgroup of G.

(5) If M_i $(i \in I)$ are left R-modules, define their **direct product** to be

$$\prod_{i \in I} M_i = \Big\{ (m_i)_{i \in I} \,\Big|\, m_i \in M_i \Big\}$$

and their **direct sum** to be

$$\bigoplus_{i\in I} M_i = \Big\{ (m_i)_{i\in I} \,\Big|\, m_i \in M_i, \text{ all but finitely many } m_i\text{'s are zero} \Big\}.$$

The module operation is defined to be

$$r \cdot (m_i)_{i\in I} = (r \cdot m_i)_{i\in I}.$$

We sometimes write the elements in $\bigoplus_{i\in I} M_i$ as $\sum_{i\in I} m_i$ instead. When this is a finite sum (i.e. $I = \{1, \dots, n\}$), we write $M_1 \oplus \cdots \oplus M_n$ instead.

(6) If $\phi : S \to R$ is a ring homomorphism, then we may naturally view an R-module M as an S-module by

$$s \cdot m = \phi(s) \cdot m.$$

Definition 13.1.4. Let R be a ring and let M and N be left R-modules.

(1) An **R-modules homomorphism** is a map $\phi : M \to N$ satisfying
- for all $x, y \in M$, $\phi(x + y) = \phi(x) + \phi(y)$, and
- for any $r \in R$ and any $x \in M$, we have $r\phi(x) = \phi(rx)$.

Write $\mathrm{Hom}_R(M, N)$ for the set of R-module homomorphisms from M to N.

(2) We say that ϕ is an **R-module isomorphism** if it is a homomorphism and a bijection.

(3) If $\phi : M \to N$ is an R-module homomorphism, then
- $\ker\phi := \{m \in M, | \, \phi(m) = 0\}$ is the kernel of ϕ; it is an R-submodule of M,
- $\mathrm{Im}(\phi) := \phi(M)$ is the image of ϕ; it is an R-submodule of N.

We have ϕ is injective if and only if $\ker\phi = \{0\}$ (or just write $\ker\phi = 0$).

As an example, $\mathbb{Z}$-module homomorphisms are the same as homomorphisms between abelian groups.

(4) If $N \subseteq M$ is an R-submodule, we define the **quotient R-module** to be

$$M/N := \{m + N \,|\, m \in M\}$$

and the operation is given by $r \cdot (m + N) = rm + N$.

(5) An R-module M is called **irreducible**, if the only R-submodule of M is 0 and M itself.

Similarly, we have isomorphism theorems for R-modules, which we leave the statements to the readers. But we mention an analogue of Jordan–Hölder theorem for modules here.

Theorem 13.1.5 (Jordan–Hölder theorem for modules)**.** *Let R be a ring and M an R-module. Assume that we are given two chains of R-submodules*

$$0 = A_0 \subseteq A_1 \subseteq \cdots \subseteq A_m = M \quad \textit{and} \quad 0 = B_0 \subseteq B_1 \subseteq \cdots \subseteq B_n = N.$$

Then setting $A'_{ij} := A_{i-1} + (A_i \cap B_j)$ and $B'_{ij} := B_{j-1} + (A_i \cap B_j)$, we may refine the above two chains by A'_{ij} and B'_{ij}, and that

$$A'_{ij}/A'_{i,j-1} \cong B'_{ij}/B'_{i-1,j}.$$

In case each subquotients A_i/A_{i-1} and B_j/B_{j-1} are irreducible, there exists a permutation σ of $\{1, \dots, n\}$ such that $A_i/A_{i-1} = B_{\sigma(i)}/B_{\sigma(i)-1}$.

13.1.6. *Homomorphisms of R-modules.* When R is a commutative ring, the set of homomorphisms between two R-modules M and N itself admits a structure of R-modules. Explicitly, for $\phi \in \mathrm{Hom}_R(M,N)$ and $a \in R$, we define $a \cdot \phi \in \mathrm{Hom}_R(M,N)$ by

$$(a \cdot \phi)(m) := a \cdot \phi(m) = \phi(a \cdot m).$$

One easily checks the standard axioms of R-module structures on $\mathrm{Hom}_R(M,N)$.

Caveat: If R is not commutative, $\mathrm{Hom}_R(M,N)$ is only an abelian group.

13.2. **Finitely generated modules.**

Definition 13.2.1. Let M be a left R-module and $X \subseteq M$ a subset. Define

$$RX := \big\{a_1x_1 + \cdots + a_nx_n \,\big|\, \text{for some } n \in \mathbb{N},\ a_1,\dots,a_n \in R,\ x_1,\dots,x_n \in X\big\}$$

to be the **submodule of M generated by X**. When $X = \{x_1,\dots,x_n\}$, we have

$$RX = \mathrm{Image}\left(\begin{array}{ccc} \phi : R^{\oplus n} & \longrightarrow & M \\ (a_1,\dots,a_n) & \longmapsto & a_1x_1+\cdots+a_nx_n. \end{array}\right)$$

(1) We call a submodule $N \subseteq M$ (possibly $N = M$) is **finitely generated** if there exists a finite set X such that $N = RX$.
(2) We call a submodule $N \subseteq M$ (possibly $N = M$) is **cyclic** if there exists $m \in M$ such that $N = Rm = \{am \,|\, a \in R\}$.
(3) We say that M is **free of rank** n if there exists $x_1,\dots,x_n \in M$ such that every $m \in M$ can be written uniquely as $m = a_1x_1 + \cdots + a_nx_n$ with $a_1,\dots,a_n \in R$.
   This is equivalent to $M \simeq R^{\oplus n}$.

Remark 13.2.2. If an R-module M is finitely generated by $x_1,\dots,x_n \in M$, then we get a surjective R-module homomorphism

$$\begin{array}{ccc} \phi : R^{\oplus n} & \longrightarrow\!\!\!\!\rightarrow & M \\ (a_1,\dots,a_n) & \longmapsto & a_1x_1+\cdots+a_nx_n. \end{array}$$

We are interested in "relations among $x_1,\dots,x_n$", e.g. elements $b_1,\dots,b_n \in R$ such that $b_1x_1 + \cdots + b_nx_n = 0$, or equivalently $\ker\phi$.

To specify an R-module homomorphism $\psi : M \to L$ to some R-module L, it is enough to specify the image $\psi(x_i)$ for each $x_i \in X$, subject to the condition that $b_1\psi(x_1) + \cdots + b_n\psi(x_n) = 0$ for each relation above.

(In nice situations, all relations are formed by taking R-linear combination of finitely many such relations; we say that M is **finitely presented** in this case. This is equivalent to that $\ker\phi$ above is also finitely generated.)

Example 13.2.3. We explain a key example of the theorem we prove later for classification of finitely generated modules over a PID.

Let F be a field and V a vector space, equipped with an action by an F-linear operator T. We define an $F[x]$-module structure on V by

$$(a_0 + a_1x + \cdots + a_nx^n)\cdot v = a_0 \cdot v + a_1x \cdot v + \cdots + a_nx^n \cdot v := a_0v + a_1T(v) + \cdots + a_nT^n(v).$$

Conversely, if V is an $F[x]$-module, we may view V as an F-module (through $F \subseteq F[x]$) so an F-vector space, then the action of x is an F-linear operator on V: for $a \in F$ and $v \in V$, we have

$$a \cdot x(v) = (ax) \cdot v = x \cdot a \cdot v = x \cdot (av).$$

In other words, we just explained a bijection:

$$\big\{F[x]\text{-modules } V\big\} \longleftrightarrow \big\{F\text{-vector spaces } V \text{ with an } F\text{-linear operator } T\big\}.$$

Similarly, an $F[x]$-submodule W of V corresponds to an F-vector space stable under the T-action.

If the vector space V is finite dimensional, then V is a finitely generated $F[x]$-module (but not necessarily conversely).

13.3. **Classification of finitely generated R-modules where R is a PID.** In this subsection, R is a PID. (The approach in this subsection looks a little tricky, but it seems to be a very "quick" approach.)

Lemma 13.3.1. *Let R be an integral domain and let N be a free R-module of rank n. Then any $n+1$ elements $x_1, \dots, x_{n+1} \in N$ are linearly dependent, i.e. there exist $a_1, \dots, a_{n+1} \in R$, not all zero, such that $a_1x_1 + \cdots + a_{n+1}x_{n+1} = 0$.*

Proof. We may identify N with $R^{\oplus n}$. Let F denote the fraction field of R. Viewing each x_i as an element of $R^{\oplus n} \subseteq F^{\oplus n}$ or even a column vector, it is well known that any $n+1$ such column vectors are linearly dependent over F, i.e. there exists $b_1, \dots, b_{n+1} \in F$ such that $b_1x_1 + \cdots + b_{n+1}x_{n+1} = 0$. Writing each b_i as c_i/d_i with $c_i, d_i \in R$ (just choose one such presentation), and put $d = d_1 \cdots d_n \in R$. Then multiply the linear relation above by d gives the needed linear relation among $x_1, \dots, x_{n+1}$. □

Theorem 13.3.2. *Let R be a PID. Let N be a free R-module of rank n and L a submodule of N. Then*

1. *L is free of rank ℓ (with $\ell \leq n$).*
2. *There exists a basis $y_1, y_2, \dots, y_n$ of N so that $a_1y_1, \dots, a_\ell y_\ell$ is a basis of L and $a_1, \dots, a_\ell \in R\backslash\{0\}$ satisfying $a_1|a_2|\cdots|a_\ell$.*

Example 13.3.3. Before giving the proof, we give an example: $R = \mathbb{Z}$, $N = \mathbb{Z}^{\oplus 3}$ and $L = \mathbb{Z}\langle(8,0,4),(12,6,0)\rangle$. If we choose the basis of N to be $e_1 = (2,0,1)$, $e_2 = (2,1,0)$, and $e_3 = (1,0,0)$. Then $L = 4\mathbb{Z}e_1 \oplus 6\mathbb{Z}e_2$. But this does not satisfy the condition $a_1|a_2$ of the theorem.

So we need to modify the above basis. If we put $e_1' = \frac{1}{2}(e_2 + e_3) = (10,3,2)$, $e_2' = \frac{1}{12}(3e_2 + 2e_3) = (4,1,1)$, and $e_3' = (1,0,0)$, we have $L = 2\mathbb{Z}e_1' \oplus 12\mathbb{Z}e_2'$. The point here is that we need to make a linear combination of the original coordinates, or rather choose a new R-module homomorphisms to take new coordinates.

Theorem 13.3.2. If $L = 0$, the theorem is trivial. Now we assume that $L \neq 0$. Our intention is to run an induction on the rank of L, but the actual argument is slightly more complicated. See later.

We first determine the value of a_1. The basic idea is that we identify N with $R^{\oplus n}$ and hence write each $x \in L$ as $(x_1, \dots, x_n)$. Our goal is to find the "minimal possible" $\gcd(x_1, \dots, x_n)$.

Fix an isomorphism $N \cong R^{\oplus n}$. Write $\pi_i : N \to R$ for the homomorphism of taking the ith coordinate. For each homomorphism $\phi : N \to R$, the image $\phi(L)$ is an R-submodule of

R, and thus an ideal. (So in particular, $\pi_i(L)$ is the set of all ith coordinates of elements of N; but we also want to allow "linear combinations" of these π_i's.)

Consider the set $\{\phi(L) \,|\, \text{any homomorphism } \phi : N \to R\}$. It contains a maximal possible ideal (a_1), say for $\phi_1 : N \to R$ with $\phi_1(y) = a_1$ for some $y \in L$. Note that $(a_1) \neq (0)$ because there exists $x \in L$ with nonzero coordinates.

We claim that, for any R-module homomorphism $\phi' : N \to R$, we have $a_1 | \phi'(y)$. Otherwise, say $d = \gcd(a_1, \phi'(y))$ then $d = r_1a_1 + r_2\phi'(y) = (r_1\phi + r_2\phi')(y)$, contradicting the maximality of (a_1).

Apply this claim to each of $\pi_1, \dots, \pi_n$, we deduce that $\pi_i(y) = a_1b_i$ for some $b_i \in R$. In other words, $y = (a_1b_1, \dots, a_1b_n)$ (namely all coordinates of y are divisible by a_1). Now we observe

$$a_1\phi(b_1, \dots, b_n) = \phi(a_1b_1, \dots, a_1b_n) = \phi(y) = a_1.$$

It follows that $\phi(b_1, \dots, b_n) = 1$. We set $y_1 := (b_1, \dots, b_n) \in N$.

Claim: (i) $N \cong Ry_1 \oplus \ker\phi$.

(ii) $L = Ra_1y_1 \oplus (L \cap \ker\phi)$ (compatibly with the direct sum decomposition of N).

Proof of (i): This is a very important typical type of argument. We are in the following situation:

$$\begin{tikzcd} N \arrow[r, two heads, "\phi"] & R. \arrow[l, bend left, "\cdot y_1"] \end{tikzcd}$$

(The following argument is general: whenever we have the above diagram with $\phi(y_1) = 1$, (1) holds.) For any $x \in N$, we may write

$$x = \underbrace{\phi(x) \cdot y_1}_{\text{multiple of } y_1} + \underbrace{(x - \phi(x)y_1)}_{\text{belongs to } \ker\phi}.$$

(To verify that $x - \phi(x)y_1 \in \ker\phi$, we compute directly that $\phi\big(x - \phi(x)y_1\big) = \phi(x) - \phi(x)\phi(y_1) = 0$.) Yet $Ry_1 \cap \ker\phi = \{ry_1 \,|\, \phi(ry_1) = r = 0\} = 0$.

Proof of (ii): For any $x \in L$, we write

$$x = \underbrace{\phi(x) \cdot y_1}_{\phi(x) \in (a_1)} + \underbrace{(x - \phi(x)y_1)}_{\text{belongs to } L \cap \ker\phi}$$

Here note that $\phi(x)y_1 \in Ra_1y_1 \subseteq L$, so $x - \phi(x)y_1 \in L$.

Thus, we may effectively reduce the discussion to $L \cap \ker\phi$.

The actual proof is a little twisted here. We need to first prove (1). Indeed, Claim (ii) implies that we have $L = Ra_1y_1 \oplus (L \cap \ker\phi)$. Applying the same discussion to $L_1 := \cap \ker\phi \subseteq N$, we see that either $L_1 = 0$ (in which case we stop the process) or $L_1 = a_2y_2 \oplus L_2$ for some other R-submodule L_2. We need to show that this process stops before getting to $L = Ra_1y_1 \oplus Ra_2y_2 \oplus \cdots \oplus Ra_{n+1}y_{n+1}$. Indeed, this is guaranteed by Lemma 13.3.1 above that the process stops at rank at most n. This proves (1).

Now we turn to prove (2), by induction on L, with a small subtlety that we reduce $L \subseteq N$ to the submodule $L \cap \ker\phi \subseteq \ker\phi$. By (1), $\ker\phi$ is still a free R-module; so the induction may proceed, and we get

$$\begin{array}{lcl} L & = & Ra_1y_1 \oplus Ra_2y_2 \oplus \cdots \oplus Ra_\ell y_\ell \\ \bigcap & & \\ N & = & Ry_1 \oplus Ry_2 \oplus \cdots \oplus Ry_n. \end{array}$$

It remains to check that $a_1|a_2$. Suppose not, $d = \gcd(a_1, a_2) = a_1r_1 + a_2r_2$. We consider the R-module homomorphism

$$\begin{aligned} \phi' : M &\longrightarrow R \\ y_1 &\longmapsto r_1 \\ y_2 &\longmapsto r_2 \\ \text{other } y_i &\longmapsto 0. \end{aligned}$$

This implies that $\phi'(a_1y_1 + a_2y_2) = a_1r_1 + a_2r_2 = d$ contradicting to the "minimality" of $\phi(L)$. So $a_1|a_2$. □

Theorem 13.3.4 (Fundamental Theorem of finitely generated modules over PIDs)**.** *Let R be a PID and L a finitely generated R-module. Then*

$$L \simeq R^{\oplus r} \oplus R/(a_1) \oplus \cdots \oplus R/(a_m) \tag{13.3.4.1}$$

with each $a_i \in R$ and $a_1|a_2|\cdots|a_m$.

Moreover, such r, $a_1, \dots, a_m$ are unique (up to associates). In particular, L is torsion-free if and only if L is free. (Here torsion-free means that for any nonzero element $x \in L$, $ax \neq 0$ for any $a \in R$.)

Corollary 13.3.5. *Classification of finitely generated abelian groups. (See Theorem 4.4.1 for the statement.)*

Proof of Theorem 13.3.4. As L is finitely generated, say by elements $x_1, \dots, x_n$, there exists a surjective homomorphism

$$\begin{aligned} \phi : R^{\oplus n} &\xrightarrow{\qquad\qquad} \!\!\!\!\rightarrow M \\ (a_1, \dots, a_n) &\longmapsto a_1x_1 + \cdots + a_nx_n. \end{aligned}$$

Then $\ker\phi \subseteq R^{\oplus n}$ is a submodule.

Applying Theorem 13.3.2 to the submodule $\ker\phi$ of the free module $R^{\oplus n}$, we see that there exists a "new basis" $y_1, \dots, y_n$ of $R^{\oplus n}$ such that $\ker\phi = \langle a_1y_1, \dots, a_my_m\rangle$. Then

$$M \cong \frac{Ry_1 \oplus \cdots \oplus Ry_n}{Ra_1y_1 \oplus \cdots \oplus Ra_my_m} \cong R/(a_1)y_1 \oplus \cdots R/(a_m)y_m \oplus R^{\oplus n-m}.$$

To show the uniqueness of (13.3.4.1), we need a different (but equivalent) form of the classification theorem. Recall that for $a \in R$ (nonzero and nonunit), if it factors as $a = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ in R with each p_i irreducible, then $R/(a) = R/(p_1^{\alpha_1}) \times \cdots \times R/(p_r^{\alpha_r})$ by Chinese remainder theorem. Thus any finitely generated R-module M can be written as

$$M \simeq R^{\oplus r} \oplus R/(p_1^{\alpha_{1,1}}) \oplus \cdots \oplus R/(p_1^{\alpha_{1,s_1}}) \oplus R/(p_2^{\alpha_{2,1}}) \oplus \cdots \tag{13.3.5.1}$$

(Conversely, given (13.3.5.1), we may resemble them to (13.3.4.1) with

$$\begin{aligned} a_1 &= p_1^{\max\{\alpha_{1,1},\dots,\alpha_{1,s_1}\}} p_2^{\max\{\alpha_{2,1},\dots,\alpha_{2,s_2}\}} \cdots \\ a_2 &= p_1^{\text{2nd largest in } \{\alpha_{1,1},\dots,\alpha_{1,s_1}\}} p_2^{\text{2nd largest in } \{\alpha_{2,1},\dots,\alpha_{2,s_2}\}} \cdots \qquad) \end{aligned}$$

We will now prove $r, p_1, p_2, \dots, \alpha_{1,1}, \dots, \alpha_{1,s_1}, \alpha_{2,1}, \dots$ are unique. We first prove a lemma.

Lemma 13.3.6. *Let p and q be prime elements such that $(p) \neq (q)$ and $m, n \in \mathbb{N}$. For an R-module M, we put $p^mM := \mathrm{Image}(M \xrightarrow{p^m\cdot} M)$.*

(1) *If $M = R$, then $p^m M \simeq p^m R$ and $p^m M/p^{m+1}M \simeq p^m M/p^{m+1}M \cong R/(p)$.*

(2) *If $M = R/(p^n)$, then*

$$p^m M \simeq \begin{cases} p^m R/p^n R & \text{if } m < n \\ 0 & \text{if } m \geq n \end{cases} \quad \text{and} \quad p^m M/p^{m+1}M \simeq \begin{cases} p^m R/p^{m+1}R \cong R/(p) & \text{if } m < n \\ 0 & \text{if } m \geq n. \end{cases}$$

(3) *If $M = R/(q^n)$, then $p^m M = 0$ and $p^m M/p^{m+1}M = 0$.*

Proof. We leave this as an exercise. For (3), we note that $(p^m, q^n) = (1)$. Indeed, as $(p, q) = (1)$, we have $1 = ap + bq$ and thus

$$1 = (ap + bq)^{m+n-1} = p^m \cdot (*) + q^n \cdot (*) \in (p^m, q^n).$$

□

Using this lemma, we see that for each prime p (say $p = p_1$) and every $m \in \mathbb{N}$, $p^m M/p^{m+1}M$ is a vector space over $R/(p)$. More precisely,

$$\begin{aligned} \dim_{R/(p)}(M/pM) &= r + \#\{\alpha_{1,1}, \ldots, \alpha_{1,s_1}\}, \\ \dim_{R/(p)}(pM/p^2M) &= r + \#\{\alpha_{1,j} \mid \alpha_{1,j} \geq 2\}, \\ \dim_{R/(p)}(p^2M/p^3M) &= r + \#\{\alpha_{1,j} \mid \alpha_{1,j} \geq 3\}, \\ \cdots & \quad \cdots \end{aligned}$$

Letting p varying over all prime elements, and all $m \in \mathbb{Z}_{\geq 0}$, we may determine all the numbers $r, p_1, p_2, \ldots, \alpha_{1,1}, \ldots$. □

14. Field extensions

14.1. **Characteristic of a field.** To study fields, we take the following viewpoint:

- start with the *prime fields*, namely $\mathbb{F}_p$ and $\mathbb{Q}$, the smallest possible fields,
- then build new fields from the known ones, e.g. $\mathbb{Q}(i)$, $\mathbb{Q}(\alpha)$ for a root of an irreducible polynomial α, or $\mathbb{Q}(x) = \mathrm{Frac}(\mathbb{Q}[x])$.

Definition 14.1.1. The **characteristic** of a field F, denoted by $\mathrm{char}(F)$, is

- the smallest positive integer p, such that $\underbrace{1 + \cdots + 1}_{p} = 0$ if such p exists, and
- 0, otherwise.

The former case is sometimes called **positive characteristic** case.

Remark 14.1.2. (1) If $\mathrm{char}(F) > 0$, it must be a prime number p. This is because if $\mathrm{char}(F) = m \cdot n$ for $m, n \in \mathbb{N}$, we have $mn = 0$ in F, and thus either $m = 0$ or $n = 0$ in F.

(2) We sometimes also use the letter K to denote field, this comes from the German word for field: Körper (body).

Definition 14.1.3. The **prime field** of a field F is the smallest field of F containing 1_F; it is

- $\mathbb{F}_p$, if $\mathrm{char}(F) = p > 0$, or
- $\mathbb{Q}$, if $\mathrm{char}(F) = 0$.

14.2. **Field extensions.**

Notation 14.2.1. If $F \subseteq K$ is a subfield of a field, we say that K is a **field extension** of F. Sometimes, we call F the **base field**.

Any field that sits as $F \subseteq E \subseteq K$ is called **intermediate fields**. We often write $K/E/F$ (reads K over E and over F), or draw as follows

$$\begin{tikzcd}
K \arrow[d, dash] \\
E \arrow[d, dash] \\
F
\end{tikzcd}$$

Note also that $F \subseteq K$ makes K an F-vector space.

Definition 14.2.2. The **degree** of the field extension K of F is $[K : F] = \dim_F K$. The extension is **finite/infinite** if $[K : F]$ is.

Theorem 14.2.3. *Let $F \subseteq E \subseteq K$ be field extensions. Then $[K : F] = [K : E][E : F]$.*

Proof. We only prove this when both extensions are finite; the infinite case can be proved similarly.

$$\begin{tikzcd}
K \arrow[d, dash] \arrow[d, dash, bend left, "{[K:E]=m}"] \arrow[dd, dash, bend right=60, "{[K:F]}"'] \\
E \arrow[d, dash] \arrow[d, dash, bend left, "{[E:F]=n}"] \\
F
\end{tikzcd}$$

Set $[K : E] = m$ and $[E : F] = n$. Let $\{\alpha_1, \dots, \alpha_m\}$ be a E-basis of K and $\{\beta_1, \dots, \beta_n\}$ be an F-basis of E. Then every element x of E can be written as a sum

$$c_1\alpha_1 + \cdots + c_m\alpha_m \quad \text{with each } c_i \in E,$$

and in turn each c_i can be written as a sum

$$c_i = d_{i1}\beta_1 + \cdots + d_{in}\beta_n \quad \text{with each } d_{ij} \in F.$$

Thus x can be written as

$$x = \sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}\alpha_i\beta_j.$$

This shows that $\{\alpha_i\beta_j \,|\, i = 1, \dots, m,\ j = 1, \dots, n\}$ generate E as an F-vector space.

Next we show that these $\alpha_i\beta_j$'s are F-linearly independent. Indeed, suppose that there exists $d_{ij} \in F$ for $i = 1, \dots, m$ and $j = 1, \dots, n$ such that

$$\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij}\alpha_i\beta_j = 0.$$

Then note that for each fixed i, $\sum\limits_{j=1}^{n} d_{ij}\beta_j \in E$. As $\alpha_1, \dots, \alpha_m$ form an E-basis of K, we must have

$$\text{for every } i, \text{ the coefficient of } \alpha_i, \text{ namely } \sum_{j=1}^{n} d_{ij}\beta_j = 0.$$

Moreover, as $\beta_1, \dots, \beta_n$ form an F-basis of E, we deduce that all $d_{ij} = 0$. □

Remark 14.2.4. A more condensed writing of the proof is:

$$K = \bigoplus_{i=1}^{m} E\alpha_i = \bigoplus_{i=1}^{m}\Big(\bigoplus_{j=1}^{n} F\beta_j\Big)\alpha_i = \bigoplus_{i=1}^{m}\bigoplus_{j=1}^{n} F\alpha_i\beta_j.$$

Example 14.2.5.

$$\begin{tikzcd}
\mathbb{Q}(\sqrt[6]{2}) = \big\{a_0 + a_1\sqrt[6]{2} + \cdots + a_5 2^{5/6} \,\big|\, a_i \in \mathbb{Q}\big\} \arrow[d, dash] \arrow[dd, dash, bend right=60, "6"'] & \\
\mathbb{Q}(\sqrt{2}) \arrow[d, dash] \arrow[d, dash, bend left, "2"] & \Rightarrow\ \big[\mathbb{Q}(\sqrt[6]{2}) : \mathbb{Q}(\sqrt{2})\big] = 3. \\
\mathbb{Q} &
\end{tikzcd}$$

14.3. **Construction of field extensions.** We start with an important fact.

Lemma 14.3.1. *Let F and E be fields. A homomorphism $\phi : F \to E$ must be injective. This then realizes E as an extension of $\phi(F) \simeq F$.*

Proof. As $\ker \phi$ is an ideal of F, namely $\{0\}$ or F. But our convention of ring homomorphisms require to send 1_F to $1_E \neq 0_E$. So $\ker \phi \neq F$, and thus $\ker \phi = \{0\}$, i.e. ϕ is injective. □

14.3.2. *Construction of field extensions.* Let F be a field and let $p(x) \in F[x]$ be an irreducible polynomial of degree n. Then $(p(x))$ is a prime ideal and hence a maximal ideal (as $F[x]$ is a PID). This implies that

$$K := F[x]/(p(x)) \quad \text{is a field.}$$

We put $\theta := x \bmod (p(x)) \in K$. Then

$$K = \big\{a_0 + a_1\theta + \cdots + a_{n-1}\theta^{n-1} \,\big|\, a_0, \ldots, a_{n-1} \in F\big\}.$$

(This follows from the fact that every polynomial $a(x)$ can be uniquely written as $a(x) = q(x)p(x) + r(x)$ with $\deg r(x) < n$; then $a(x) \bmod (p(x)) = r(x)$.)

So K is an F-vector space of dimension n. Moreover, F embeds in K as constant polynomials.

We say that K is the **extension of F of degree n determined by** $p(x)$.

Lemma 14.3.3. *Equation $p(x) = 0$ has a zero in K.*

Proof. Assume that $p(x) = p_0 + p_1x + \cdots + p_nx^n$. Then

$$p(\theta) = p_0 + p_1\theta + \cdots + p_n\theta^n = p_0 + p_1x + \cdots + p_nx^n + (p(x)) = 0 + (p(x)).$$

So θ is a "tautological" zero of $p(x) = 0$ in K. □

Example 14.3.4. (1) Recall the natural isomorphism

$$\begin{tikzcd} \mathbb{R}[x]/(x^2+1) \arrow[r, "\simeq"] & \mathbb{C} \\ a + bx \arrow[r, mapsto] & a + b\mathbf{i} \end{tikzcd}$$

But from now on, we will try to distinguish these two: $\mathbb{R}[x]/(x^2+1)$ is an abstractly constructed field extension of $\mathbb{R}$. It is isomorphic to $\mathbb{C}$. BUT there are TWO WAYS of to make such an isomorphism

$$\phi_1,\, \phi_2 : \mathbb{R}[x]/(x^2+1) \xrightarrow{\simeq} \mathbb{C},$$

$$\phi_1(a + bx) = a + b\mathbf{i} \quad \text{and} \quad \phi_2(a + bx) = a - b\mathbf{i}.$$

(2) For $K = \mathbb{Q}[x]/(x^3 - 2)$, we have three realizations:

- Realization 1: given by

$$\begin{tikzcd} \iota_1 : K \arrow[r, hook] & \mathbb{R} \\ x \arrow[r, mapsto] & \sqrt[3]{2}. \end{tikzcd}$$

So $K \simeq \iota_1(K)$ is a subfield of $\mathbb{R}$.

- Realizations 2 and 3: given by

$$\begin{aligned} \iota_2, \iota_3 : K &\hookrightarrow \mathbb{R} \\ x &\longmapsto \begin{cases} \iota_2(x) = e^{2\pi\mathbf{i}/3}\sqrt[3]{2} \\ \iota_3(x) = e^{4\pi\mathbf{i}/3}\sqrt[3]{2} \end{cases} \end{aligned}$$

  So $\iota_2(K)$ and $\iota_3(K)$ are different fields from $\iota_1(K)$. But they are abstractly isomorphic (for the purpose of algebraic operations).

(3) $\mathbb{F}_2[x]/(x^2+x+1)$ is a field extension of $\mathbb{F}_2$ of degree 2. This gives K a field of 4 elements.

Definition 14.3.5. Let K be an extension of F, and let $\alpha_1, \dots, \alpha_n \in K$.

(1) The **field extension of** F **generated by** $\alpha_1, \dots, \alpha_n$, denoted by $F(\alpha_1, \dots, \alpha_n)$, is the smallest subfield of K containing F.
(2) If $K = F(\alpha)$ for some $\alpha \in K$, then we say that K is a **simple extension** of F.
(3) If $K = F(\alpha_1, \dots, \alpha_n)$, we say that K is a **finitely generated extension** of F.

We remark that $F(\alpha_1, \alpha_2) = (F(\alpha_1))(\alpha_2)$.

Theorem 14.3.6. *Let K be a field extension of F and let $\alpha \in K$. We have a dichotomy:*

(1) *either $1, \alpha, \alpha^2, \dots$ are linearly independent over F, in which case $F(\alpha) \simeq F(x) = \mathrm{Frac}(F[x])$,*
(2) *or $1, \alpha, \alpha^2, \dots$ are linearly dependent over F, in which case, there exists a unique monic polynomial $m_\alpha(x) = m_{\alpha,F}(x)$, called the* **minimal polynomial of** α **over** F*, that is irreducible over F and $m_\alpha(\alpha) = 0$.*
  Moreover, $F(\alpha) = F[x]/(m_\alpha(x))$ and $[F(\alpha) : F] = \deg m_\alpha(x)$.

Proof. Consider case (1): the condition implies that

$$\begin{aligned} \phi : F[x] &\hookrightarrow K \\ f(x) &\longmapsto f(\alpha) \end{aligned}$$

is an injective homomorphism. This clearly extends to a homomorphism

$$\begin{aligned} \phi : F(x) &\hookrightarrow K \\ f(x)/g(x) &\longmapsto f(\alpha)/g(\alpha) \end{aligned}$$

as $g(\alpha) \neq 0$. This ϕ must be injective by Lemma 14.3.1. Its image is $\phi(F(x)) = F(\alpha)$.

Consider case (2): In this case,

$$\begin{aligned} \phi : F[x] &\longrightarrow K \\ f(x) &\longmapsto f(\alpha) \end{aligned}$$

is *not* injective. Then $\ker\phi = (p(x))$ is a prime ideal and hence a maximal ideal. We may take $p(x)$ to be monic.

This $p(x)$ is the minimal polynomial of α, as it is the nonzero polynomial with minimal degree in $\ker\phi$. Thus

$$F(\alpha) = \mathrm{Im}(\phi) \simeq F[x]/(p(x)).$$

□

Definition 14.3.7. Keep the setup in the above theorem.

- In case (1), we call α **transcendental over** F.
- In case (2), we call α **algebraic over** F.

We say that the extension K over F is **algebraic** if every element α of K is algebraic over F (or equivalently $[F(\alpha) : F]$ is finite).

14.4. Finite versus algebraic extensions.

Example 14.4.1. A typical algebraic yet not finite field extension is $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \dots)$ over $\mathbb{Q}$.

Theorem 14.4.2. *The following are equivalent for a field extension K of F:*

(1) *K is a finite extension of F.*
(2) *K is finitely generated and algebraic over F.*

Proof. (1) $\Rightarrow$ (2). If K is a finite extension of F, K is generated over F by the basis element (of K over F). For any $\alpha \in K$, $[F(\alpha) : F] \leq [K : F]$ is finite; so α is algebraic over F.

(2) $\Rightarrow$ (1). We will prove this after some preparation. □

Corollary 14.4.3. *If K is a finite extension of F and $\alpha \in K$, then $[F(\alpha) : F] \,\big|\, [K : F]$.*

Example 14.4.4. If K/F is a field extension of prime degree, then any element $\alpha \in K$ that is not in F generates K over F.

Lemma 14.4.5. *Given field extensions of $K/E/F$ and $\alpha \in K$, then*

$$m_{\alpha,E}(x) \big| m_{\alpha,F}(x)$$

as polynomials in $E[x]$. In particular,

$$\deg(m_{\alpha,E}(x)) \leq \deg(m_{\alpha,F}(x)).$$

Proof. This is because $m_{\alpha,F}(\alpha) = 0$. So viewing this in the polynomial ring $E[x]$, we have $m_{\alpha,F}(x) \in (m_{\alpha,E}(x))$. This implies that $m_{\alpha,E}(x) \,|\, m_{\alpha,F}(x)$ in $E[x]$ and thus $\deg(m_{\alpha,E}(x)) \leq \deg(m_{\alpha,F}(x))$. □

Corollary 14.4.6. *Given field extensions of $K/E/F$ and $\alpha \in K$, then*

$$[E(\alpha) : E] \leq [F(\alpha) : F].$$

Proof. This is because $[E(\alpha) : E] = \deg m_{\alpha,E}(x)$ and $[F(\alpha) : F] = \deg m_{\alpha,F}(x)$. □

Definition 14.4.7. Let K be a finite extension of F and $F \subseteq K_i \subseteq K$ for intermediate fields K_1 and K_2. Define the **composite** of K_1 and K_2 to be

$$K_1K_2 := \text{minimal field that contains both } K_1 \text{ and } K_2.$$

Example 14.4.8. Inside $\mathbb{C}$, the composite of $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ is

$$\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \big\{a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6} \,\big|\, a, b, c, d \in \mathbb{Q}\big\}.$$

Corollary 14.4.9. *Let K_1 and K_2 be two intermediate fields in the field extension K over F such that $[K_i : F] < +\infty$. Then*

$$[K_1K_2 : F] \leq [K_1 : F] \cdot [K_2 : F].$$

Proof. As K_1 is a finite extension of F, we may write $K_1 = F(\alpha_1, \ldots, \alpha_n)$. We consider the following tower

$$\begin{tikzcd}
 & K_1K_2 \arrow[d, dash] \\
K_1 \arrow[ur, dash] \arrow[d, dash] & \vdots \arrow[d, dash] \\
\vdots \arrow[d, dash] & K_2(\alpha_1, \alpha_2) \arrow[d, dash] \\
F(\alpha_1, \alpha_2) \arrow[ur, dash] \arrow[d, dash] & K_2(\alpha_1) \arrow[d, dash] \\
F(\alpha_1) \arrow[ur, dash] \arrow[d, dash] & K_2 \\
F \arrow[ur, dash] &
\end{tikzcd}$$

Applying Corollary 14.4.6 to each parallelogram, we have

$$\begin{aligned}
[K_2(\alpha_1) : K_2] &\leq [F(\alpha_1) : F] \\
[K_2(\alpha_1, \alpha_2) : K_2(\alpha_1)] &\leq [F(\alpha_1, \alpha_2) : F(\alpha_1)] \\
\cdots \quad & \quad \cdots
\end{aligned}$$

Taking the product of these inequalities gives $[K_1K_2 : K_2] \leq [K_1 : F]$. This implies the inequality of this corollary. $\square$

14.4.10. *Continued with the proof of Theorem 14.4.2.*

(2) ⇒ (1). Assume that K is finitely generated algebraic extension of F. We may then write $K = F(\alpha_1, \ldots, \alpha_n) = F(\alpha_1)F(\alpha_2)\cdots F(\alpha_n)$ with each α_i algebraic over F. Then Corollary 14.4.9 implies that

$$[K : F] \leq [F(\alpha_1) : F] \cdots [F(\alpha_n) : F]$$

is finite. $\square$

Corollary 14.4.11. *Let K be a field extension of F and let $\alpha, \beta \in K$ be elements algebraic over F. Then $\alpha \pm \beta$, $\alpha\beta$, and α/β (when $\beta \neq 0$) are all algebraic over F.*

Proof. This is because $\alpha \pm \beta$, $\alpha\beta$, and α/β all belong to the field $F(\alpha, \beta)$ which is a finite extension of F. $\square$

Definition 14.4.12. Let K be a field extension of F. It follows from the above corollary that

$$\big\{\alpha \in K \,|\, \alpha \text{ is algebraic over } F\big\}$$

is a subfield of K, called the **algebraic closure of F in K**.

Example 14.4.13. Consider the field extension $\mathbb{C}$ of $\mathbb{Q}$. We put

$$\overline{\mathbb{Q}} := \big\{\alpha \in \mathbb{C} \,\big|\, \alpha \text{ is algebraic over } \mathbb{Q}\big\}$$

the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$. Note that the condition α being algebraic over $\mathbb{Q}$ is equivalent to the condition that α is a zero of a monic polynomial $f(x) \in \mathbb{Q}[x]$.

Theorem 14.4.14. *If L/K and K/F are both algebraic extensions, then L/F is algebraic.*

Proof. Let $\alpha \in L$. Its minimal polynomial $m_\alpha(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ over K involves only finitely many elements of K, each of them being algebraic over F. So we see that $F(\alpha)$ is contained in the field extension

$$F(a_0, a_1, \ldots, a_{n-1})(\alpha)$$

over F, which is finite. So L is algebraic over F. □

15. Normal extensions

15.1. **Splitting fields.**

Definition 15.1.1. Given a field F and a polynomial $f(x) \in F[x]$ of degree n (not necessarily irreducible), a field extension K of F is called a **splitting field** of $f(x)$ over F if

(1) $f(x)$ splits completely in $K[x]$: $f(x) = (x - \alpha_1) \cdots (x - \alpha_n)$ for $\alpha_1, \ldots, \alpha_n \in K$, and
(2) $K = F(\alpha_1, \ldots, \alpha_n)$.

Remark 15.1.2. If K is a splitting field of $f(x) \in F[x]$ over F and E is an intermediate field of K over F, then K is a splitting field of $f(x) \in E[x]$ over E.

Theorem 15.1.3. *For any field F and $f(x) \in F[x]$ of degree n, a splitting field K of F exists. Moreover, $[K : F] \le n!$.*

Proof. We use induction on $\deg f(x) = n$. When $n = 1$, F itself is the splitting field of $f(x)$ over F. Suppose that the statement is proved for polynomials of strictly smaller degrees (over *any* field).

Let $p(x)$ be an irreducible factor of $f(x)$. Then

$$L := F[x]/(p(x))$$

is a field extension of F of degree $\deg p(x) \le \deg f(x) = n$, over which $p(x)$ has a zero. This implies that

$$f(x) = (x - \theta) \cdot g(x).$$

By inductive hypothesis, $g(x)$ over L admits a splitting field K of degree $[K : L] \le (\deg g)! = (n-1)!$; this K is also a splitting field of $f(x)$ over F. Thus,

$$[K : F] = [K : L][L : F] \le (n-1)! \cdot \deg p(x) \le n!.$$

□

15.2. **Uniqueness of splitting fields.**

Lemma 15.2.1. *If $\eta : F \xrightarrow{\simeq} F'$ is an isomorphism of fields and $p(x) \in F[x]$ is irreducible, then $p'(x) := \eta(p(x)) \in F'[x]$ is irreducible, and η induces a natural isomorphism*

$$\eta : F[x]/(p(x)) \xrightarrow{\simeq} F'[x]/(p'(x)).$$

Remark 15.2.2. In abstract algebra, we usually do not use $p'(x)$ to denote the derivative of a polynomial. Typically, $p'(x)$ is just *another* polynomial.

Example 15.2.3. Lemma 15.2.1 seems to be trivial, but it can be used to prove the following seemingly nontrivial statement. Consider the isomorphism

$$\begin{aligned} \eta : \mathbb{Q}(\sqrt{2}) &\xrightarrow{\ \simeq\ } \mathbb{Q}(\sqrt{2}) \\ a + b\sqrt{2} &\longmapsto a - b\sqrt{2}. \end{aligned}$$

Then we have a natural isomorphism

$$\mathbb{Q}\big(\sqrt{5+\sqrt{2}}\big) \simeq \mathbb{Q}(\sqrt{2})[x]/(x^2 - 5 - \sqrt{2}) \xrightarrow{\simeq} \mathbb{Q}(\sqrt{2})[x]/(x^2 - 5 + \sqrt{2}) \simeq \mathbb{Q}\big(\sqrt{5-\sqrt{2}}\big).$$

Proposition 15.2.4. *Let $\eta : F \xrightarrow{\simeq} F'$ be an isomorphism of fields and $f(x) \in F[x]$. Put $f'(x) := \eta(f(x)) \in F'[x]$. If E is a splitting field of $f(x)$ over F and E' is a splitting field of $f'(x)$ over F', then there exists a (not necessarily unique) isomorphism $\sigma : E \xrightarrow{\simeq} E'$ restricting to $\eta : F \xrightarrow{\simeq} F'$, i.e.*

$$\begin{tikzcd}
E \arrow[r, "\sigma", "\simeq"'] & E' \\
\bigcup & \bigcup \\
F \arrow[r, "\eta", "\simeq"'] & F'.
\end{tikzcd}$$

Proof. We will prove that for the splitting field K of $f(x)$ over F constructed in Theorem 15.1.3, we have the following commutative diagram with top row being isomorphisms.

$$\begin{tikzcd}
E & K \arrow[l, "\sigma_1"', "\simeq"] \arrow[r, "\sigma'", "\simeq"'] & E' \\
\bigcup & \bigcup & \bigcup \\
F \arrow[r, equal] & F \arrow[r, "\eta", "\simeq"'] & F'.
\end{tikzcd}$$

(The upshot here is that we use an explicitly constructed splitting field to aid the construction of σ.) We will construct the right diagram and the left diagram can be constructed similarly. In fact, we will prove a slightly stronger statement.

Claim: If $\eta : F \xrightarrow{\simeq} F'$ is an isomorphism of fields and E' an extension of F' over which $\eta(f(x))$ splits completely, then η extends to a homomorphism $\sigma' : K \to E'$, making the following diagram commute.

$$\begin{tikzcd}
K \arrow[r, dashed, "\sigma'"] & E' \\
\bigcup & \bigcup \\
F \arrow[r, "\eta", "\simeq"'] & F'.
\end{tikzcd}$$

The proposition follows immediately from the claim.

Proof of the claim: As in the proof of Theorem 15.1.3, we make an induction on $\deg(f)$. At each step, we consider the following diagram

$$\begin{tikzcd}
L = F[x]/(p(x)) \arrow[r, dashed, "??"] \arrow[d, dash] & E' \\
F \arrow[ur, hook, "\eta"'] &
\end{tikzcd}$$

where $\eta : F \simeq F' \subseteq E'$ is the given embedding. As $\eta(p(x)) = p'(x)$ has a zero in E', say $\alpha \in E'$, we may define a homomorphism

$$\begin{aligned}
\sigma'_L : L = F[x]/(p(x)) &\longrightarrow E' \\
x + (p(x)) &\longmapsto \alpha.
\end{aligned}$$

By induction, we eventually get an embedding $\sigma' : K \hookrightarrow E'$ compatible with $\sigma'_L : L \to E'$ and hence with $F \simeq F'$. This proves the claim and the proposition. □

Lemma 15.2.5. *If we have the following tower of field extensions*

$$\begin{tikzcd}
 & K & \\
E \arrow[ur, dash] & & E' \arrow[ul, dash] \\
 & F \arrow[ul, dash] \arrow[ur, dash] &
\end{tikzcd}$$

such that both E and E' are splitting fields of some polynomial $f(x) \in F[x]$. Then $E = E'$ (as an equality of subfields of K).

Proof. By the definition of splitting field, we may

- split $f(x)$ over E as $c(x - \alpha_1) \cdots (x - \alpha_n)$ with $c \in E^\times$ and $\alpha_1, \dots, \alpha_n \in E$, and
- split $f(x)$ over E' as $c(x - \alpha'_1) \cdots (x - \alpha'_n)$ with $c' \in E'^\times$ and $\alpha'_1, \dots, \alpha'_n \in E'$.

But viewing these two factorizations in $K[x]$, we must have $\{\alpha_1, \dots, \alpha_n\} = \{\alpha'_1, \dots, \alpha'_n\}$. This forces $E = F(\alpha_1, \dots, \alpha_n) = F(\alpha'_1, \dots, \alpha'_n) = E'$. □

Lemma 15.2.6. *Consider a tower of extensions $K/E/F$. If E is a splitting field over F of some polynomial $f(x) \in F[x]$, then for any automorphism $\sigma : K \xrightarrow{\simeq} K$ such that $\sigma|_F = \mathrm{id}$, we must have $\sigma(E) = E$.*

Proof. This is because $\sigma(E)$ is a splitting field of $\sigma(f) = f$. By the above lemma, we deduce that $\sigma(E) = E$. □

15.3. **Intrinsic definition of splitting fields.**

Definition 15.3.1. An algebraic extension K of F is called **normal** if

- for any irreducible polynomial $f(x) \in F[x]$ that has one zero in K, $f(x)$ splits completely over K.

This definition of normal extension appears to be a very strong condition at first sight. Yet we have the following.

Theorem 15.3.2. *A finite extension K of F is normal if and only if it is the splitting field of some $f(x) \in F[x]$.*

Proof. We first prove that a finite normal extension K of F is a splitting field. Indeed, write $K = F(\alpha_1, \dots, \alpha_r)$ for $\alpha_1, \dots, \alpha_r \in K$. For each α_i, the minimal polynomial $m_{\alpha_i}(x) \in F[x]$ splits over K by normality. Thus, K is the splitting field of $m_{\alpha_1}(x) \cdots m_{\alpha_r}(x)$ over F.

Conversely, we assume that K is a splitting field of $f(x) \in F[x]$ over F. We aim to prove that K is a normal extension of F.

If $p(x) \in F[x]$ is an irreducible polynomial that has a zero $\alpha \in K$, let L be the splitting field of $p(x)$ over K. We want to prove that $L = K$.

$$\begin{tikzcd}
L \arrow[d, dash, "\text{splitting field of } p(x)"'] \\
K \arrow[d, dash, "\text{splitting field of } f(x)"'] \\
F
\end{tikzcd}$$

Clearly, L is the splitting field of $f(x)p(x)$ over F.

Let β be another zero of $p(x)$ in L. Then there exists an isomorphism

$$\begin{aligned} \eta : F(\alpha) &\xrightarrow{\simeq} F(\beta) \\ \alpha &\longmapsto \beta \end{aligned}$$

fixing the field F. By Theorem 15.2.4, this isomorphism extends to an isomorphism $\sigma : L \simeq L$ (as L is the splitting field of $f(x)p(x)$ over $F(\alpha)$ and $F(\beta)$).

But by Lemma 15.2.6, we have $\sigma(K) = K$. Yet $\alpha \in K$, we must have $\beta = \eta(\alpha) = \sigma(\alpha) \in K$. This means that all zeros of $p(x)$ belong to K, or equivalently, $p(x)$ splits completely over K. So K is a normal extension of F. □

Corollary 15.3.3. *If K is a finite and normal extension of F, for any intermediate field E, the field K is a normal extension of E. (Note that E need not be a normal extension of F.)*

Proof. By the theorem above, K is a splitting field of some $f(x) \in F[x]$ over F. Hence K is a splitting field of $f(x)$ over E. The reverse implication of the above theorem implies that K is normal over E. □

Exercise 15.3.4. Remove the finiteness hypothesis on the extension K/F in the previous corollary. Namely only assume that K is a normal extension of F and E an intermediate field, prove that K is normal over E.

Definition 15.3.5. If K is an algebraic extension of F, a **normal closure** of K over F is a field extension L of K such that

(1) L is a normal extension of F, and
(2) L is the minimal such extension, i.e. if $L \supseteq L' \supseteq K$ is so that L'/F is normal, then $L = L'$.

In particular, if K is finite over F, the normal closure L is also finite over F.

Lemma 15.3.6. *A normal closure of a finite extension K over F exists and is unique up to (some) isomorphism.*

Proof. Existence: Assume that $K = F(\alpha_1, \dots, \alpha_r)$. Put

$$f(x) := \prod_{i=1}^{r} m_{\alpha_i,F}(x) \in F[x].$$

Take L to be a splitting field of $f(x)$ over K. This implies that L is a splitting field over F, showing that L is normal over F and is a normal closure of K over F.

Uniqueness: If L' is another normal closure of K over F, then by the Claim in Proposition 15.2.4, there exists an embedding $L \hookrightarrow L'$. But then $f(x)$ already splits over L. This implies by minimality that $L = L'$. □

Remark 15.3.7. (1) The isomorphism between two different normal closures is not unique (this can be understood by later discussion of Galois groups).

(2) An infinite normal extension K/F is an increasing union of its normal subextensions that finite over F. So by taking certain limit (and using an axiom of choice), the above lemma implies that for any algebraic extension K/F, the normal closure exists and is unique up to some isomorphism.

Example 15.3.8. (1) The splitting field of $x^2 - 2$ is $\mathbb{Q}(\sqrt{2})$.

(2) The splitting field of $x^3 - 2$ is

$$\begin{tikzcd}
 & \mathbb{Q}(\sqrt[3]{2}, \zeta_3) \arrow[ddl, dash] \arrow[d, dash] \arrow[dr, dash] \arrow[drr, dash] & & \\
 & \mathbb{Q}(\sqrt[3]{2}) \arrow[dd, dash] & \mathbb{Q}(\zeta_3\sqrt[3]{2}) \arrow[ddl, dash] & \mathbb{Q}(\zeta_3^2\sqrt[3]{2}) \arrow[ddll, dash] \\
\mathbb{Q}(\zeta_3) \arrow[dr, dash] & & & \\
 & \mathbb{Q}. & &
\end{tikzcd}$$

Here and later, for $n \in \mathbb{N}$, $\zeta_n = e^{2\pi \mathbf{i}/n}$ is a primitive nth root of unity.

(3) The splitting field of $x^n - 1 = \prod_{i=0}^{n-1}(x - \zeta_n^i)$ is $\mathbb{Q}(\zeta_n)$. (We call $\mathbb{Q}(\zeta_n)$ the nth cyclotomic field.) We will see later that $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$.

For example, $x^p - 1 = (x-1)(x^{p-1} + \cdots + x + 1)$, where the latter factor is irreducible. So $[\mathbb{Q}(\zeta_p) : \mathbb{Q}] = p - 1$.

(4) The splitting field of $x^p - t$ over $\mathbb{F}_p(t)$ is $\mathbb{F}_p(t^{1/p})$:

$$\begin{tikzcd}
L = \mathbb{F}_p(t^{1/p}) \arrow[d, dash] \\
\mathbb{F}_p(t)
\end{tikzcd}$$

Over L, we have a factorization

$$x^p - t = (x - t^{1/p})^p.$$

This situation is very "strange" because the irreducible polynomial factors into a polynomial where roots have multiplicity.

15.4. **Perfect fields.** We investigate the pathology appearing in Example 15.3.8(4). Let F be a field. If $\mathrm{char}(F) = 0$, we will show that there is no pathology in the next lecture. For now, we only discuss the positive characteristic situation.

Definition 15.4.1. If the characteristic of a field F is a prime number p, define the **Frobenius endomorphism** of F to be

$$\sigma = \sigma_F : F \to F, \qquad \sigma(x) = x^p.$$

(Check: for $x, y \in F$, $\sigma(xy) = \sigma(x)\sigma(y)$ and

$$\sigma(x+y) = (x+y)^p = x^p + px^{p-1}y + \binom{p}{2}x^{p-2}y^2 + \cdots + y^p = x^p + y^p = \sigma(x) + \sigma(y),$$

where we note that all the terms in the middle of the expression are multiples of p, which is zero in F.)

We say that F is a **perfect** field if σ is an isomorphism, otherwise, we say that F is **imperfect**. This is equivalent to say that every element $a \in F$ is a pth power (of a unique element in F).

Example 15.4.2. (1) $F = \mathbb{F}_p$ is a perfect field (so is any finite field). This is because $\sigma : F \to F$ is always injective, so also surjective by counting elements.

(2) $F = \mathbb{F}_p(t)$ is *not* a perfect field. Explicitly, $\sigma(\mathbb{F}_p(t)) = \mathbb{F}_p(t^p)$ is a proper subfield of $\mathbb{F}_p(t)$.

(3) $F = \mathbb{F}_p(t, t^{1/p}, t^{1/p^n}; n \in \mathbb{N})$ is a perfect field.

Proposition 15.4.3. *Algebraic extensions of perfect fields are still perfect.*

Proof. Let K be an algebraic extension of F with F a perfect field of characteristic p. It suffices to show that each $\alpha \in K$ admits a pth root in $E := F(\alpha)$ (E is a finite extension of F).

Consider the Frobenius endomorphism $\sigma : E \to E$ given by $a \mapsto a^p$ (which is compatible with the Frobenius endomorphism on F). We view the images of the Frobenius endormophisms as subfields $\sigma(E)$ and $\sigma(F)$ of E and F, respectively, i.e. we have the following field extensions:

$$\begin{tikzcd}
& E \arrow[dd, dash] \\
\sigma(E) \arrow[ur, dash] \arrow[dd, dash] & \\
& F \\
\sigma(F) \arrow[ur, dash] &
\end{tikzcd}$$

This implies that

$$[E : \sigma(E)] \cdot [\sigma(E) : \sigma(F)] = [E : F] \cdot [F : \sigma(F)].$$

As σ induces isomorphisms $F \xrightarrow{\sigma} \sigma(F)$ and $E \xrightarrow{\sigma} \sigma(E)$, we must have $[E : F] = [\sigma(E) : \sigma(F)]$. Plugging this back to the equation above gives that $[E : \sigma(E)] = [F : \sigma(F)]$ which is 1 as F is perfect. It follows that $\sigma(E) = E$ and E is perfect. □

Remark 15.4.4. Carefully inspecting the above proposition, we see that, if one defines, for a field F, the imperfect degree $t \in \mathbb{N}$ by $[F : \sigma(F)] = p^t$, then if E is a *finite* extension of F, then E and F have the same imperfect degree. (We will see in the next lecture that $F/\sigma(F)$ is purely inseparable; so the degree is a power of p.)

However, imperfect degree may not be preserved under infinite algebraic extension. For example $\mathbb{F}_p(t)$ has imperfect degree 1 yet $\mathbb{F}_p(t, t^{1/p^n}; n \in \mathbb{N})$ is perfect.

Remark 15.4.5. For a field F of characteristic $p > 0$, we may define its **perfection**:

$$F^{\mathrm{perf}} := \bigcup_n \sigma^{-n}(F).$$

For example, the perfection of $\mathbb{F}_p(t)$ is just $\mathbb{F}_p(t, t^{1/p^n}; n \in \mathbb{N})$. Moreover, F is perfect if and only if $F = F^{\mathrm{perf}}$.

16. Separable extensions and finite fields

16.1. **Separable polynomials.** Recall from the previous lecture that a field F of characteristic $p > 0$ is perfect if the Frobenius map $\sigma : F \to F$, $\sigma(a) = a^p$ is an isomorphism.

Definition 16.1.1. If F is a field and $f(x) = a_0 + a_1x + \cdots + a_nx^n \in F[x]$ is a polynomial, define its **formal derivative** to be

$$D(f)(x) := a_1 + 2a_2x + \cdots + na_nx^{n-1} \in F[x].$$

If $f(x) = c(x - \alpha_1)^{e_1} \cdots (x - \alpha_r)^{e_r} \in F[x]$ with α_i pairwise distinct, we say that α_i is a zero of $f(x)$ with multiplicity e_i.

Theorem 16.1.2. *Let $f(x) \in F[x]$ be a polynomial of degree ≥ 1. Then $f(x)$ has no repeated roots in its splitting field K if and only if $(f(x), D(f)(x)) = (1)$.*

Proof. "$\Leftarrow$" If $(f(x), D(f)(x)) = (1)$, we have

$$f(x)p(x) + D(f)(x)q(x) = 1$$

for some polynomials $p(x), q(x) \in F[x] \subseteq K[x]$. But if $(x - \alpha)^2 \,|\, f(x)$ for some $\alpha \in K$, we have $(x - \alpha) \,|\, D(f)(x)$. So the above equality implies that $(x - \alpha) \,|\, 1$. This is absurd! So $f(x)$ has no repeated root in K.

"$\Rightarrow$" Assume that $f(x) = (x - \alpha_1) \cdots (x - \alpha_n)$ for $\alpha_i \in K$ is the factorization of $f(x)$ in $K[x]$. Assume that $(d(x)) = (f(x), D(f)(x))$ in $F[x]$ with $d(x)$ monic. Since the multiplicity of each zero of $f(x)$ is one, we must have

$$D(f)(\alpha_i) \neq 0.$$

So $d(x) \,|\, D(f)(x)$ implies that $x - \alpha_i$ cannot divide $d(x)$. Yet $d(x) \,|\, f(x) = (x - \alpha_1) \cdots (x - \alpha_n)$. So $d(x) = 1$. □

Corollary 16.1.3. *If $f(x)$ is an irreducible polynomial in $F[x]$, then $f(x)$ has repeated roots in its splitting field if and only if $D(f)(x) = 0$.*

Proof. By Theorem 16.1.2, the polynomial $f(x)$ has repeated zero if and only if $(f(x), D(f)(x)) = (1)$. As $f(x)$ is irreducible, this is further equivalent to $f(x) \,|\, D(f)(x)$, which is in turn equivalent to $D(f)(x) = 0$ (because $D(f)$ has lower degree than $f(x)$). □

Definition 16.1.4. Let $f(x)$ be an irreducible polynomial in $F[x]$.

- If $f(x)$ has repeated roots in its splitting field (or equivalently $D(f)(x) = 0$), we say that f is **inseparable**;
- If $f(x)$ has only simple roots, we say f is **separable**.

Corollary 16.1.5. *If $\operatorname{char}(F) = 0$, all irreducible polynomials are separable.*

Proof. This is because when $f(x) \neq 0$ and $\deg f \geq 1$, we have $D(f)(x) \neq 0$. □

Corollary 16.1.6. *If $\operatorname{char}(F) = p > 0$, and if $f(x)$ is inseparable, then*

$$f(x) = g(x^p) \quad \textit{for some } g \in F[x] \textit{ irreducible.}$$

Moreover, this can only happen when F is imperfect.

Proof. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be an irreducible polynomial with repeated roots. Corollary 16.1.3 implies that

$$D(f)(x) = a_1 + 2a_2x + \cdots + na_nx^{n-1} = 0.$$

This means that $ia_i = 0$ and thus $a_i = 0$ when $p \nmid i$. So

$$f(x) = a_0 + a_px^p + \cdots + a_{2p}x^{2p} = \cdots = g(x^p) \quad \text{with } g(x) = a_0 + a_px + a_{2p}x^2 + \cdots .$$

This polynomial $g(x)$ is clearly irreducible.

Finally, we show that this cannot happen when F is perfect. Indeed, if F is perfect, then every $a_{pi} = b_i^p$ for some $b_i \in F$. Thus

$$f(x) = b_0^p + b_1^px^p + b_2^px^{2p} + \cdots = (b_0 + b_1x + b_2x^2 + \cdots)^p$$

is not irreducible, contradicting with our initial assumption. □

Corollary 16.1.7. *If* $\mathrm{char}(F) = p > 0$*, an irreducible polynomial* $f(x) \in F[x]$ *is of the form* $f(x) = g(x^{p^e})$ *with* $g(x) \in F[x]$ *is an irreducible and separable polynomial and* $e \in \mathbb{Z}_{\geq 0}$*. In this case,* $f(x)$ *has* $\deg(g)$ *distinct zeros in its splitting field.*

Proof. The first statement follows from the previous corollary. For the second statement, we note that, if $g(x) = \prod\limits_i (x - \alpha_i)$ in a splitting field of F, then we have

$$f(x) = \prod_i (x^{p^e} - \alpha_i) = \prod_i \left(x - \alpha_i^{1/p^e}\right)^{p^e}.$$

□

16.2. **Separable extensions.**

Definition 16.2.1. Let K be an algebraic extension of F.

- We say an element $\alpha \in K$ is **separable** or **inseparable** if its minimal polynomial $m_{\alpha,F}(x)$ is.
- We say that K is **separable** over F if every element $\alpha \in K$ is separable over F. Otherwise, we say that K is an inseparable extension of F.

Remark 16.2.2. If E is an intermediate field of an algebraic extension K of F, and if $\alpha \in K$ is separable over F, then α is separable over E. This is because $m_{\alpha,E}(x)$ divides $m_{\alpha,F}(x)$ in $E[x]$.

Theorem 16.2.3. (1) *If* α *is separable over* F*, then* $F(\alpha)$ *is a separable extension of* F*.*
(2) *If* K/E *and* E/F *are separable extensions, then* K/F *is separable.*

This theorem will be proved later, after we introduce some useful tools to study field extensions.

Construction 16.2.4. Let K be a finite extension of F and M a normal extension of F that contains K (e.g. M is the normal closure of K over F). We consider all possible homomorphisms $\phi : K \to M$ (which is automatically injective) such that $\phi|_F = \mathrm{id}$. We use $\mathrm{Hom}_F(K, M)$ to denote this set. Graphically,

$$\begin{tikzcd}
K \arrow[r, dashed] \arrow[d, dash] & M \\
F \arrow[ur, dash] &
\end{tikzcd}$$

Example 16.2.5. We consider the case when $F = \mathbb{Q}$, $K = \mathbb{Q}(\sqrt[3]{2})$ and $M = \mathbb{Q}(\sqrt[3]{2}, \zeta_3)$.

$$\begin{tikzcd}
 & M = \mathbb{Q}(\sqrt[3]{2}, \zeta_3) \\
K = \mathbb{Q}(\sqrt[3]{2}) \arrow[ur, dash] \arrow[d, dash] & \\
F = \mathbb{Q} &
\end{tikzcd}$$

The set $\mathrm{Hom}_{\mathbb{Q}}(K, M) = \mathrm{Hom}_{\mathbb{Q}}(\mathbb{Q}[x]/(x^3 - 2), M)$ is given by

$$\begin{aligned}
K = \mathbb{Q}(\sqrt[3]{2}) &\longrightarrow M = \mathbb{Q}(\sqrt[3]{2}, \zeta_3) \\
\varphi_0 : \sqrt[3]{2} &\longmapsto \sqrt[3]{2} \\
\varphi_1 : \sqrt[3]{2} &\longmapsto \sqrt[3]{2} \cdot \zeta_3 \\
\varphi_2 : \sqrt[3]{2} &\longmapsto \sqrt[3]{2} \cdot \zeta_3^2.
\end{aligned}$$

Note that in this example, $\#\mathrm{Hom}_F(K, M) = [K : F]$.

The following lemma generalized the above observation.

Lemma 16.2.6. *If $K = F(\alpha)$ with $m_{\alpha,F}(x) = g(x^{p^e})$ for some $g \in F[x]$ irreducible and separable, then*

$$\#\mathrm{Hom}_F(F(\alpha), M) = \deg g(x) \leq [F(\alpha) : F]$$

with equality if and only if α is separable.

Proof. Such a $\phi \in \mathrm{Hom}_F(F(\alpha), M)$ is determined by where α goes.

$$\begin{tikzcd}
K = F(\alpha) \arrow[r, dashed, "\phi"] \arrow[d, dash] & M \arrow[dl, dash] \\
F &
\end{tikzcd}$$

The constraint on $\phi(\alpha)$ is that, it must be a zero of $m_{\alpha,F}(x)$ in M. There are precisely $\deg(g)$ of them. □

Remark 16.2.7. The lemma implies that $\#\mathrm{Hom}_F(F(\alpha), M)$ does NOT depend on M, as long as it is normal over F.

Corollary 16.2.8. *Let K be a finite extension of F and M a normal extension of F containing K. Then*

$$\#\mathrm{Hom}_F(K, M) \leq [K : F]. \tag{16.2.8.1}$$

Moreover, the following are equivalent:

1. *$K = F(\alpha_1, \ldots, \alpha_n)$ with each α_i separable over F.*
2. *The equality in* (16.2.8.1) *holds.*
3. *The field extension K of F is separable, i.e. any $\alpha \in K$ is separable over F.*

Note that the equivalence of conditions in the above corollary implies Theorem 16.2.3(1).

Proof. We consider the following tower of extensions and their embeddings into M.

$$\begin{tikzcd}
K \arrow[d, dash] & \\
\vdots \arrow[d, dash] & M \\
F(\alpha_1, \alpha_2) \arrow[ur, dashed, hook] \arrow[d, dash] & \\
F(\alpha_1) \arrow[uur, dashed, hook] \arrow[d, dash] & \\
F \arrow[uuur, hook] &
\end{tikzcd}$$

By Corollary 16.2.8, we have

$$\#\mathrm{Hom}_F(F(\alpha_1), M) \leq [F(\alpha_1) : F].$$

The equality holds if and only if α_1 is separable over F.

For each embedding $F(\alpha_1) \hookrightarrow M$, we have

$$\#\mathrm{Hom}_{F(\alpha_1)}\big(F(\alpha_1, \alpha_2), M\big) \leq \big[F(\alpha_1, \alpha_2) : F(\alpha_1)\big].$$

In other words, the number of embeddings $F(\alpha_1, \alpha_2) \hookrightarrow M$ extending any embedding $F(\alpha_1) \hookrightarrow M$ is at most $[F(\alpha_1, \alpha_2), F(\alpha_1)]$. The equality holds if and only if α_2 is separable over $F(\alpha_1)$. Combining the above two inequalities gives that

$$\#\mathrm{Hom}_F\big(F(\alpha_1, \alpha_2), M\big) \leq \big[F(\alpha_1, \alpha_2) : F\big].$$

We see that, by induction, we may deduce (16.2.8.1).

Now we prove the equivalence among (1)–(3).

(3) ⇒ (1) is obvious.

(1) ⇒ (2): by Remark 16.2.2, α_i separable over F implies that α_i separable over $F(\alpha_1, \dots, \alpha_{i-1})$. By the argument above and note that each inequality above is an equality, so the equality in (16.2.8.1) holds.

(2) ⇒ (3): if some α is not separable, then $\#\mathrm{Hom}_F(F(\alpha), M) < [F(\alpha) : F]$. For each embedding $F(\alpha) \hookrightarrow M$, (16.2.8.1) implies that

$$\#\mathrm{Hom}_{F(\alpha)}(K, M) \leq [K : F(\alpha)].$$

This implies that $\#\mathrm{Hom}_F(K, M) < [K : F]$, contradicting (2). □

16.2.9. *Proof of Theorem 16.2.3(2).* We need to show that if K is a separable extension of E, and E is a separable extension of F, then K is a separable extension of F.

Without loss of generality, we may assume that K is a finite extension of F. (This in fact requires some explanation: we need to show that every $\alpha \in K$ is separable over F. Let $m_{\alpha,E}(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ be the minimal polynomial of α over E. Put $E' := F(a_0, \dots, a_{n-1})$; it is a finite separable extension over F. Then $m_{\alpha,E}(x) = m_{\alpha,E'}(x)$ is a separable polynomial. It follows from (1) that $K' := E'(\alpha)$ is a finite separable extension over K' and E' a finite separable extension of F. We are reduced to the tower of *finite* extensions $K'/E'/F$.)

Take M a normal extension of F containing K. Then

$$\#\mathrm{Hom}_F(E, M) = [E : F].$$

For each given embedding $E \hookrightarrow M$, $\#\mathrm{Hom}_E(K, M) = [K : E]$. Thus, we have

$$\#\mathrm{Hom}_F(K, M) = [K : F].$$

This follows from the equivalence relation in Corollary 16.2.8. □

In light of Theorem 16.2.3, we make the following definition.

Definition 16.2.10. If K is a finite extension of F, put

$$K^s := \{\alpha \in K \,|\, \alpha \text{ is separable over } F\};$$

it is the maximal intermediate field that is separable over F (by Theorem 16.2.3).

Define

$$[K : F]_{\text{sep}} := [K^s : F] \quad \text{and} \quad [K : F]_{\text{insep}} := [K : K^s].$$

Remark 16.2.11. It is an exercise to modify the above proof of Theorem 16.2.3 to show that, if M is a normal extension of F containing K, then

$$[K : F]_{\text{sep}} = \#\mathrm{Hom}_F(K, M),$$

and $[K : F]_{\text{insep}}$ is always a power of p.

Moreover, for a tower of finite extension $K/E/F$, we have

$$[K : F]_{\text{sep}} = [K : E]_{\text{sep}} \cdot [E : F]_{\text{sep}} \quad \text{and} \quad [K : F]_{\text{insep}} = [K : E]_{\text{insep}} \cdot [E : F]_{\text{insep}}.$$

16.3. **Primitive element theorem.**

Theorem 16.3.1. *A finite separable extension of fields is generated by one element.*

In fact, we have a stronger statement: if $K = F(\alpha, \beta)$ with α, β algebraic over F and β separable over F. Then $K = F(\gamma)$ for some $\gamma \in K$.

Example 16.3.2. A typical non-monogenic extension is $K = \mathbb{F}_p(x^{1/p}, y^{1/p})$ over $F = \mathbb{F}_p(x, y)$ of degree p^2. Indeed, for any $\alpha \in K$, $\alpha^p \in F$, so $[F(\alpha) : F] \leq p$.

Remark 16.3.3. The stronger version of the theorem in fact implies that if F is a field with imperfect degree ≤ 1, then any finite extension of F is generated by one element.

Proof of Theorem 16.3.1. The basic idea is that: most $\theta = \alpha + c\beta$ for $c \in F$ should generate K over F; we just need to avoid the "bad" c's.

We will prove the case of finite fields by a separate argument (later in this lecture). Now we assume that $\#F = \infty$.

Let $f(x)$ and $g(x)$ be minimal polynomials of α and β over F, respectively. Let E be a splitting field of $f(x)g(x)$ and $\alpha = \alpha_1, \dots, \alpha_r$, $\beta = \beta_1, \dots, \beta_s$ the distinct zeros of $f(x)$ and $g(x)$.

We take $c \in F$ such that $\alpha_i + c\beta_1 \neq \alpha_k + c\beta_j$ for any i, k as long as $j \neq 1$. (This only rules out finitely many choices of $c \in F$.)

Set $\theta := \alpha_1 + c\beta_1$. Then $F(\theta) \subseteq F(\alpha, \beta)$. We want to prove that $\alpha, \beta \in F(\theta)$ (and thus $F(\alpha, \beta) = F(\theta)$.

Consider the polynomials $f(\theta - cx)$ and $g(x)$. They have common zero (in E) if and only if for some $x = \beta_j$, there exist indices i such that $\theta - c\beta_j = \alpha_i$, i.e. when $\alpha_1 + c\beta_1 = \theta = \alpha_i + c\beta_j$. By our choice of c, the only such choice is that $x = \beta_1$. Thus, we deduce that in $F(\theta)[x]$, we have $\gcd\big(f(\theta - cx), g(x)\big)$ has only one zero.

$$\big(f(\theta - cx), g(x)\big) = (x - \beta_1)$$

in $E[x]$, and hence in $F(\theta)[x]$. This in particular implies that $\beta = \beta_1 \in F(\theta)$ and hence $\alpha = c\beta - \theta \in F(\theta)$. The theorem is proved. □

16.4. **Finite fields.**

Theorem 16.4.1. (1) *If F is a finite field, then* $\mathrm{char}(F) = p > 0$ *for a prime p, and* $\#F = p^n$ *for* $n = [F : \mathbb{F}_p]$.

(2) *For each p^n, there is a field F of p^n elements. It is a splitting field of $x^{p^n} - x \in \mathbb{F}_p[x]$, so it is unique up to isomorphisms.*

Proof. (1) is clear.

(2) If F is a finite field of p^n elements, $F^\times$ is a finite group. By Corollary 12.3.5, $F^\times$ is a cyclic group of order $p^n - 1$. This implies that for any $a \in F^\times$, $a^{p^n-1} - 1 = 0$. Thus, all elements in F are zeros of $x^{p^n} - x = 0$, and they are exactly the p^n zeros of $x^{p^n} - x$. In other words, F is the splitting field of $x^{p^n} - x$ over $\mathbb{F}_p$. (Such splitting field is unique up to isomorphisms.)

Conversely, if F is the splitting field of $x^{p^n} - x = 0$ over $\mathbb{F}_p$. Note that $D(x^{p^n} - x) = p^n x^{p^n-1} - 1 = -1$ in $\mathbb{F}_p[x]$ and thus $(x^{p^n} - x, D(x^{p^n} - x)) = (1)$ in $\mathbb{F}_p[x]$. So $x^{p^n} - x$ has only simple zeros in F. This implies that $x^{p^n} - x$ has exactly p^n zeros in F.

We claim that these p^n zeros form a subfield of F (and thus it must be equal to F). This is because whenever $\alpha, \beta \in F$ (with $\beta \neq 0$) satisfy $\alpha^{p^n} = \alpha$ and $\beta^{p^n} = \beta$, we always have $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$, and α/β are all zeros of $x^{p^n} - x$. □

Definition 16.4.2. We write $\mathbb{F}_{p^n}$ for a field of p^n elements (which is unique up to isomorphisms).

Lemma 16.4.3. (1) *The field $\mathbb{F}_{p^m}$ can be viewed as a subfield of $\mathbb{F}_{p^n}$ if and only if $m|n$. In this case, as a subset, $\mathbb{F}_{p^m} \subseteq \mathbb{F}_{p^n}$ is unique.*

(2) *The field $\mathbb{F}_{p^n} = \mathbb{F}_p(\alpha)$ for some α with* $\deg m_{\alpha,\mathbb{F}_p}(x) = n$. *In particular, finite extensions of finite fields are monogenic.*

Proof. (1) If $\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$, then $\mathbb{F}_{p^n}$ is a vector space over $\mathbb{F}_{p^m}$. Thus, p^n is a power of p^m. So $m|n$.

Conversely, if $m|n$, $\mathbb{F}_{p^m}$ is a splitting field of $x^{p^m} - x$. Yet $\mathbb{F}_{p^n}$ splits

$$x^{p^n} - x = (x^{p^m} - x) \cdot \frac{x^{p^n-1} - 1}{x^{p^m-1} - 1}.$$

Thus we have $\mathbb{F}_{p^m}$ embeds in $\mathbb{F}_{p^n}$. Or more precisely, $\mathbb{F}_{p^m}$ is the set of zeros of $x^{p^m} - x$ inside $\mathbb{F}_{p^n}$.

(2) Take any $\alpha \in \mathbb{F}_{p^n} \backslash \bigcup_{m|n,\, m\neq n} \mathbb{F}_{p^m}$. The number of such element is: if $n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$,

$$p^n \Big(1 - \frac{1}{p^{p_1}}\Big) \cdots \Big(1 - \frac{1}{p^{p_r}}\Big) > 0.$$

This implies that $[\mathbb{F}_p(\alpha) : \mathbb{F}_p] = n$ (as α is not contained in any other subfields of $\mathbb{F}_{p^n}$). So $m_{\alpha,\mathbb{F}_p}(x)$ has degree n. □

17. Galois theory I

This and the next lecture are the high point of this semester where we discuss the Galois theory of algebraic field extensions. This is a beautiful theory that describes algebraic field extensions in a nice and clean way, and relates together with the group theory we learned earlier.

17.1. **Galois group.** We first recall Corollary 16.2.8: let K be a finite extension of F and let M be a normal extension of F containing K.

$$\begin{tikzcd}
K \arrow[r, dashed, "\phi"] \arrow[d, no head] & M \arrow[dl, no head] \\
F &
\end{tikzcd}$$

Then $\#\mathrm{Hom}_F(K, M) \leq [K : F]$ and the equality holds if K is separable over F.

Definition 17.1.1. We say a algebraic extension K of F is **Galois** if it is separable and normal. (We will mostly assume that K/F is finite in this lecture and discuss the case of infinite extensions in a later lecture.)

(When K/F is finite,) we define $\mathrm{Gal}(K/F) := \mathrm{Aut}_F(K)$ to be the group of automorphisms $\phi : K \to K$ such that $\phi|_F = \mathrm{id}$. It is called the **Galois group** of K over F.

Lemma 17.1.2. *If K is a finite Galois extension of F, then $\#\mathrm{Gal}(K/F) = [K : F]$.*

Proof. Consider the above situation we recalled with $K = M$ Galois over F, then each $\phi \in \mathrm{Hom}_F(K, K)$ is an automorphism of K fixing F. This is because $\phi : K \to K$ is an injective F-linear map between the same finite dimensional F-vector space; so ϕ must be bijective. Thus, we have $\mathrm{Hom}_F(K, K) = \mathrm{Aut}_F(K)$ and

$$\#\mathrm{Hom}_F(K, K) = [K : F].$$

This means that $\#\mathrm{Gal}(K/F) = [K : F]$. □

The following will be a typical example for Galois theory to study the interplay between Galois group and intermediate fields.

Example 17.1.3. Consider $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$ as an extension of $\mathbb{Q}$. The Galois group is equal to $\mathrm{Gal}(K/\mathbb{Q}) = \{\mathrm{id}, \sigma, \tau, \sigma\tau\}$, where

- id is the identity map;
- $\sigma(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a - b\sqrt{2} + c\sqrt{3} - d\sqrt{6}$;
- $\tau(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a + b\sqrt{2} - c\sqrt{3} - d\sqrt{6}$;
- $\sigma\tau(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}) = a - b\sqrt{2} - c\sqrt{3} + d\sqrt{6}$.

We can compute the subfields that are invariant under these automorphisms:

$$\begin{aligned}
K^{\sigma=1} &= \{x \in K \,|\, \sigma(x) = x\} = \{a + c\sqrt{3} \,|\, a, c \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{3}), \\
K^{\tau=1} &= \{x \in K \,|\, \sigma(x) = x\} = \{a + b\sqrt{2} \,|\, a, b \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{2}), \\
K^{\sigma\tau=1} &= \{x \in K \,|\, \sigma(x) = x\} = \{a + d\sqrt{6} \,|\, a, d \in \mathbb{Q}\} = \mathbb{Q}(\sqrt{6}).
\end{aligned}$$

We may rearrange the field extensions by the following diagram

$$\begin{tikzcd}
 & K = \mathbb{Q}(\sqrt{2}, \sqrt{3}) \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
K^{\sigma=1} = \mathbb{Q}(\sqrt{3}) \arrow[dr, dash] & K^{\tau=1} = \mathbb{Q}(\sqrt{2}) \arrow[d, dash] & K^{\sigma\tau=1} = \mathbb{Q}(\sqrt{6}) \arrow[dl, dash] \\
 & \mathbb{Q} &
\end{tikzcd}$$

Remark 17.1.4. From this example, we can see that there is a general statement: if H is a group acting on a field K by automorphisms, then

$$K^H := \big\{x \in K \,\big|\, \sigma(x) = x \text{ for any } \sigma \in H\big\}$$

is a subfield.

Theorem 17.1.5 (Galois theory)**.** *Let K be a finite Galois extension with $G = \mathrm{Gal}(K/F)$.*

(1) *Then there is a one-to-one correspondence between*

$$\begin{array}{rcl}
\big\{\textit{subgroups } H \leq G\big\} & \longleftrightarrow & \big\{\textit{intermediate fields } E \textit{ of } K/F\big\} \\
H & \longmapsto & K^H = \{x \in K \,|\, h(x) = x \textit{ for any } h \in H\} \\
\mathrm{Gal}(K/E) = \{g \in G \,|\, g|_E = \mathrm{id}|_E\} & \longleftarrow\!\shortmid & E
\end{array}$$

(2) *The correspondence is inclusion-reversive, i.e.*

$$H_1 \subseteq H_2 \iff K^{H_1} \subseteq K^{H_2}.$$

(3) *The correspondence turns degrees of field extensions into indices of subgroups, namely, we have*

$$\#H = [K : K^H] \quad \textit{and} \quad [G : H] = [K^H : F]$$

(4) *If $H \longleftrightarrow E$, then*

$$gHg^{-1} \longleftrightarrow g(E).$$

(5) *$H \leq G$ is a normal subgroup $\iff$ K^H is a normal extension of F. In this case,*

$$\mathrm{Gal}(K^H/F) \cong G/H.$$

(6) *If $H_1, H_2 \longleftrightarrow E_1, E_2$, then*

$$H_1 \cap H_2 \longleftrightarrow E_1E_2 \quad \textit{and} \quad \langle H_1, H_2\rangle \longleftrightarrow E_1 \cap E_2.$$

The proof of part (1) is more complicated, which we leave to the next lecture. Assuming (1), we prove (2)–(6) to give the readers some idea of how Galois theory works.

Proof of Theorem 17.1.5(2)–(6).

(2) $H_1 \subseteq H_2 \Rightarrow K^{H_1} \supseteq K^{H_2}$ because if for any $x \in K$, $H_2x = x$ implies $H_1x = x$.

$E_1 \subseteq E_2 \Rightarrow \mathrm{Gal}(K/E_2) \subseteq \mathrm{Gal}(K/E_1)$ because for any $h \in \mathrm{Gal}(K/F)$, $h|_{E_2} = \mathrm{id} \Rightarrow h_{E_1} = \mathrm{id}$.

(3) $\#H = [K : K^H]$ is proved in Lemma 17.1.2 above. The other equality follows from

$$\#H \cdot [G : H] = \#G \overset{K/F \text{ Galois}}{=} [K : F] = [K : K^H] \cdot [K^H : F].$$

(4) If $E = K^H$, we need to show that $g(E) = K^{gHg^{-1}}$. This is because, for $x \in K$,

$$\begin{aligned} & ghg^{-1}(x) = x \text{ for any } h \in H \\ \Leftrightarrow\ & hg^{-1}(x) = g^{-1}(x) \text{ for any } h \in H \\ \Leftrightarrow\ & g^{-1}(x) \in E \\ \Leftrightarrow\ & x \in g(E). \end{aligned}$$

(5) We prove that $H \leq G$ is a normal subgroup $\Leftrightarrow$ K^H is a normal extension of F.
Proof of "$\Leftarrow$": if K^H is a normal extension of F,

$$\begin{tikzcd} K \arrow[d, dash] \\ K^H \arrow[d, dash] \\ F \end{tikzcd}$$

then any automorphism $\sigma : K \to K$ that fixes F must also stabilizers K^H by Lemma 15.2.6. In other words,

$$K^H = \sigma(K^H) = K^{\sigma H \sigma^{-1}} \quad \Rightarrow \quad H = \sigma H \sigma^{-1}.$$

So H is a normal subgroup of G.

Proof of "$\Rightarrow$": if $H \lhd G$ is a normal subgroup, then we need to show that K^H is a normal extension of F, namely, if $f(x)$ is an irreducible polynomial in $F[x]$ that has a zero α in K^H, then $f(x)$ splits completely over K.

For this, we need a useful **Lemma**: if K is a Galois extension of F, and $f(x)$ is an irreducible polynomial that splits over K. If α is a root of $f(x)$, then other zeros are exactly $\{\sigma(\alpha) \,|\, \sigma \in \mathrm{Gal}(K/F)\}$.

The proof of **Lemma**: Clearly, each $\sigma(\alpha)$ is a zero of $f(x)$. Write

$$g(x) := \prod_{g \in \mathrm{Gal}(K/F)} (x - \sigma(\alpha));$$

its coefficients are all invariant under $\mathrm{Gal}(K/F)$-action and thus belong to F. In $K[x]$, we know that $(f(x), g(x)) \neq (1)$, so in $F[x]$, $(f(x), g(x)) \neq (1)$. But $f(x)$ is irreducible; we must have $f(x) \,|\, g(x)$. This means that all zeros of $f(x)$ are of the form $\sigma(\alpha)$ for some $\sigma \in \mathrm{Gal}(K/F)$. The lemma is proved.

We come back to the proof of "$\Rightarrow$" of (5). The lemma implies that all zeros of $f(x)$ are $\sigma(\alpha)$'s for some $\sigma \in \mathrm{Gal}(K/F)$. Thus

$$\alpha \in K^H \ \Rightarrow \ \sigma(\alpha) \in K^{\sigma H \sigma^{-1}} = K^H$$

This means that $f(x)$ splits over K^H.

When both conditions of (5) are satisfied, each element $g \in G$ fixes K^H and thus defines an automorphism in $\mathrm{Gal}(K^H/F)$. This defines a group homomorphism $\pi : G \to \mathrm{Gal}(K^H/F)$. The kernel of π is precisely

$$\ker \pi = \{\sigma : K \to K \mid \sigma|_{K^H} = \mathrm{id}\} = \mathrm{Gal}(K/K^H) = H.$$

By first isomorphism theorem of homomorphism, we have an embedding $\bar{\pi} : G/H \hookrightarrow \mathrm{Gal}(K^H/F)$. There are two ways to see that $\bar{\pi}$ is surjective.

Method 1: by counting, $\#G/H \overset{(3)}{=} [K^H : F] = \#\mathrm{Gal}(K^H/F)$, $\bar{\pi}$ must be an isomorphism.
Method 2: essentially by Proposition 15.2.4, any isomorphism $K^H \to K^H$ extends to an isomorphism $K \to K$. So π is surjective.

In any case, we have shown that $\mathrm{Gal}(K^H/F) \cong G/H$.

(6) If H_1 and H_2 correspond to intermediate fields E_1 and E_2, respectively. Then we have

$$\begin{aligned}\mathrm{Gal}(K/E_1E_2) &= \{h \in \mathrm{Gal}(K/F) \,|\, h|_{E_1E_2} = \mathrm{id}\} \\ &= \{h \in \mathrm{Gal}(K/F) \,|\, h|_{E_1} = \mathrm{id} \text{ and } h|_{E_2} = \mathrm{id}\} = H_1 \cap H_2.\end{aligned}$$

$$K^{\langle H_1, H_2\rangle} = K^{H_1} \cap K^{H_2} = E_1 \cap E_2.$$

□

Remark 17.1.6. One thing we observe during the proof of Theorem 17.1.5(2)–(6) is that, we often encounter the situation where for some results, it is easier to start on the group side and deduce results on the field extension side, but some other statement, it is easier to start from the field side. The Galois correspondence allows us to always start from the easy side. For a concrete example, when proving (6), suppose that we want to show that $K^{H_1 \cap H_2} = K^{H_1}K^{H_2}$ directly, it would be very difficult.

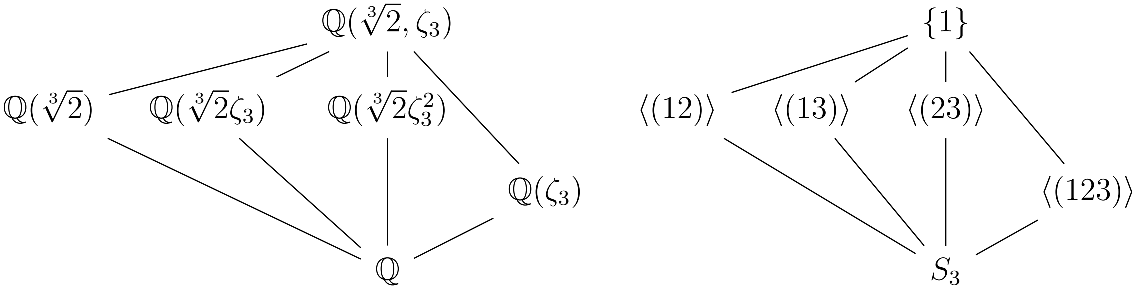
Remark 17.1.7. We have encountered earlier multiple times of non-canonical isomorphisms. For example, two normal closures L and L' of a finite separable extension K/F are isomorphic but there is no canonical isomorphism. This can be seen as follows: if $\eta : L \to L'$ is an isomorphism that fixes K. Note that L is finite separable and normal over F, so it is Galois. Then $\mathrm{Gal}(L/K)$ is a subgroup of $\mathrm{Gal}(L/F)$. For any $h \in \mathrm{Gal}(L/K)$, we may modify the isomorphism $\eta : L \cong L'$ to $L \xrightarrow{h} L \xrightarrow{\eta} L'$. This gives rise to another isomorphism between L and L' that is identity on K. In fact, all isomorphisms between L and L' that are identity on K arise this way. Yet there is no distinguished choice among these isomorphisms.

17.2. **Examples of Galois extensions.**

Example 17.2.1. Corresponding to Example 17.1.3, we have the following corresponding diagram

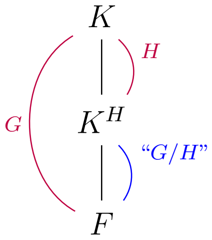
$$\begin{tikzcd}
 & \mathbb{Q}(\sqrt{2},\sqrt{3}) \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\mathbb{Q}(\sqrt{2}) \arrow[dr, dash] & \mathbb{Q}(\sqrt{3}) \arrow[d, dash] & \mathbb{Q}(\sqrt{6}) \arrow[dl, dash] \\
 & \mathbb{Q} &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \{1\} \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & & \\
\{1,\tau\} \arrow[dr, dash] & \{1,\sigma\} \arrow[d, dash] & \{1,\sigma\tau\} \arrow[dl, dash] & \\
 & \{1,\sigma,\tau,\sigma\tau\} & = & \mathbf{Z}_2 \times \mathbf{Z}_2
\end{tikzcd}$$

Example 17.2.2.



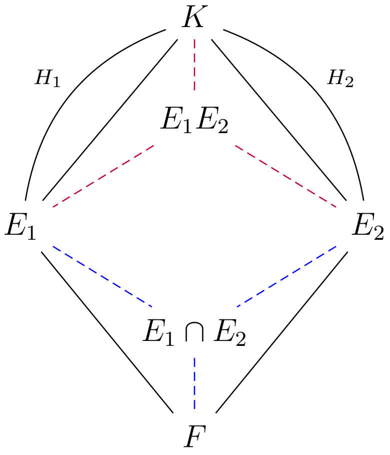
Note that $\langle(123)\rangle$ is a normal subgroup; this corresponds to that $\mathbb{Q}(\zeta_3)$ is a Galois extension of $\mathbb{Q}$.

Remark 17.2.3. An alternative way to represent the field extensions and Galois correspondences is the following:



The advantage of such notation is that, in a self-explanatory way, $\#G = [K : F]$ and $\#H = [K : E]$. When E is normal over F (or equivalently when H is a normal subgroup of G, the Galois group of E/F is the quotient G/H.

Moreover, Theorem 17.1.5(6) can be graphically expressed (not proved) as



Graphic picture tells us that $\mathrm{Gal}(K/E_1E_2)$ should be contained in both H_1 and H_2 and should be the maximal subgroup of G with that property; so $\mathrm{Gal}(K/E_1K_2) = H_1 \cap H_2$. Similarly, $\mathrm{Gal}(K/E_1 \cap E_2)$ should contain both H_1 and H_2 and should be the minimal subgroup of G with that property, i.e. $\mathrm{Gal}(K/E_1 \cap E_2) = \langle H_1, H_2\rangle$.

Example 17.2.4. We give a case of Galois group being D_8, with $\theta = \sqrt[4]{2}$. The normal closure of $\mathbb{Q}(\sqrt[4]{2})$ is $\mathbb{Q}(\sqrt[4]{2}, \mathbf{i})$. There are obvious automorphisms of $\mathbb{Q}(\sqrt[4]{2})$ given by

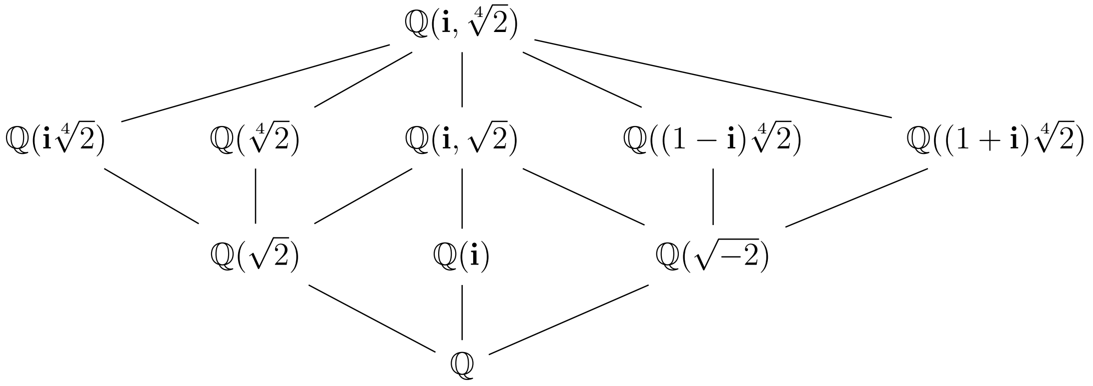
$$\begin{aligned} s(\mathbf{i}) &= -\mathbf{i}, \qquad & s(\sqrt[4]{2}) &= \sqrt[4]{2}, \\ r(\sqrt[4]{2}) &= \mathbf{i}\sqrt[4]{2}, \qquad & r(\mathbf{i}) &= \mathbf{i}. \end{aligned}$$

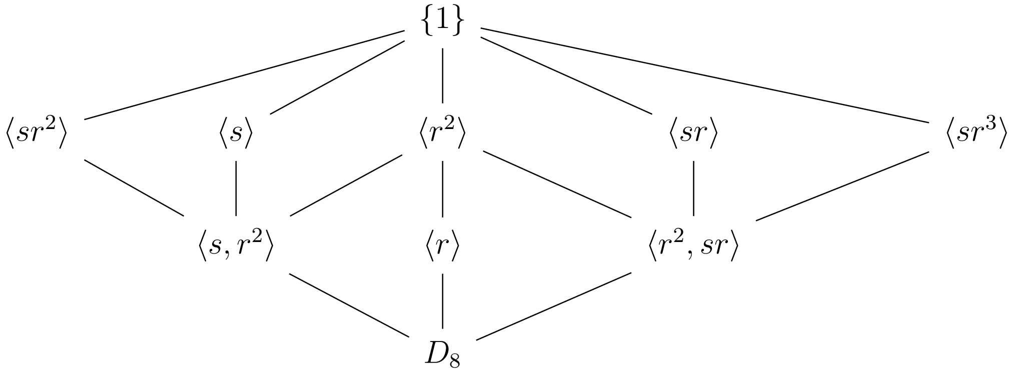
One can check that

$$\begin{aligned} rsrs &: \sqrt[4]{2} \xmapsto{\quad s \quad} \sqrt[4]{2} \xmapsto{\quad r \quad} \mathbf{i}\sqrt[4]{2} \xmapsto{\quad s \quad} -\mathbf{i}\sqrt[4]{2} \xmapsto{\quad r \quad} \sqrt[4]{2}, \\ rsrs &: \mathbf{i} \xmapsto{\quad s \quad} -\mathbf{i} \xmapsto{\quad r \quad} -\mathbf{i} \xmapsto{\quad s \quad} \mathbf{i} \xmapsto{\quad r \quad} \mathbf{i}. \end{aligned}$$

We have $srs = r^{-1}$ and thus

$$D_8 = \langle r, s \mid r^4 = s^2 = 1,\ srs = r^{-1} \rangle.$$


$\mathbb{Q}(\mathbf{i}, \sqrt[4]{2})$
$\mathbb{Q}(\mathbf{i}\sqrt[4]{2})$
$\mathbb{Q}(\sqrt[4]{2})$
$\mathbb{Q}(\mathbf{i}, \sqrt{2})$
$\mathbb{Q}((1-\mathbf{i})\sqrt[4]{2})$
$\mathbb{Q}((1+\mathbf{i})\sqrt[4]{2})$
$\mathbb{Q}(\sqrt{2})$
$\mathbb{Q}(\mathbf{i})$
$\mathbb{Q}(\sqrt{-2})$
$\mathbb{Q}$



Here, it is quite straightforward to obtain the left part if the diagram. To get the invariants under $\langle sr \rangle$, we simply take the element $\sqrt[4]{2}$ and consider its "trace" under the group action of $\langle sr \rangle$ and $\langle sr^3 \rangle$, respectively:

$$\begin{aligned} \sqrt[4]{2} + sr(\sqrt[4]{2}) &= \sqrt[4]{2} - \mathbf{i}\sqrt[4]{2} = (1-\mathbf{i})\sqrt[4]{2}, \\ \sqrt[4]{2} + sr^3(\sqrt[4]{2}) &= \sqrt[4]{2} + \mathbf{i}\sqrt[4]{2} = (1+\mathbf{i})\sqrt[4]{2}. \end{aligned}$$

These are certainly elements contained in the corresponding invariant fields. On the other hand, we can check the minimal polynomials of $(1-\mathbf{i})\sqrt[4]{2}$ and $(1+\mathbf{i})\sqrt[4]{2}$ are both equal to $x^4 - 8$, so we are now clear about these fields.

Remark 17.2.5. In general, if $\theta = \sqrt[n]{a}$ with $n \in \mathbb{N}$ and $a \in \mathbb{Q}$ such that for any divisor m of n, $\sqrt[m]{a} \notin \mathbb{Q}$, then the normal closure of $\mathbb{Q}(\theta)$ is $K = \mathbb{Q}(\theta, \zeta_n)$ (exercise!). The Galois group $\mathrm{Gal}(K/\mathbb{Q})$ is always a *subgroup* of semidirect product $\mathbf{Z}_n \rtimes \mathbf{Z}_n^\times$.[5]

In "majority" cases, the Galois group $\mathrm{Gal}(K/\mathbb{Q})$ is isomorphic to $\mathbf{Z}_n \rtimes \mathbf{Z}_n^\times$. A partial picture of the intermediate fields is

$$\begin{tikzcd}
 & \mathbb{Q}(\sqrt[n]{a}, \zeta_n) \arrow[dd, dash] \\
\mathbb{Q}(\sqrt[n]{a}) \arrow[ur, dash] \arrow[dd, dash] & \\
 & \mathbb{Q}(\zeta_n) \\
\mathbb{Q} \arrow[ur, dash] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \{1\} \arrow[dd, dash] \\
\mathbf{Z}_n^\times \arrow[ur, dash] \arrow[dd, dash] & \\
 & \mathbf{Z}_n \\
\mathbf{Z}_n \rtimes \mathbf{Z}_n^\times \arrow[ur, dash] &
\end{tikzcd}$$

Note that $\mathbf{Z}_n$ is a normal subgroup of $\mathbf{Z}_n \rtimes \mathbf{Z}_n^\times$, which corresponds to that $\mathbb{Q}(\zeta_n)$ is a Galois extension of $\mathbb{Q}$. On the other hand, $\mathbf{Z}_n^\times$ is a not a normal subgroup and thus $\mathbb{Q}(\sqrt[n]{a})$ is not normal over $\mathbb{Q}$.

Corollary 17.2.6. *If K is a Galois extension of F with Galois group $G = \mathrm{Gal}(K/F)$ being an abelian group, then any intermediate field E is Galois over F.*

Proof. This is because any subgroup of an abelian group is normal. □

17.3. **Finite fields.** Let $q = p^r$ be a power of p. We have shown that there exists a unique finite field of q elements: $\mathbb{F}_q$. It is a perfect field.

We have shown that, for each $n \in \mathbb{N}$, there is a finite extension $\mathbb{F}_{q^n}$ of $\mathbb{F}_q$, consisting of q^n elements; it is a normal and separable extension of $\mathbb{F}_q$.

$$\begin{tikzcd}
\mathbb{F}_{q^n} \arrow[d, dash] \\
\mathbb{F}_q
\end{tikzcd}$$

Definition 17.3.1. The **q-Frobenius automorphism** of $\mathbb{F}_{q^n}$ is the automorphism

$$\begin{aligned} \phi_q : \mathbb{F}_{q^n} &\longrightarrow \mathbb{F}_{q^n} \\ a &\longmapsto a^q \end{aligned}$$

(It is the same as composing the p-Frobenius r times.)

Lemma 17.3.2. *The Galois group of $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$ is isomorphic to $\mathbf{Z}_n$, generated by ϕ_q.*

Proof. For the q-Frobenius automorphism ϕ_q of $\mathbb{F}_{q^n}$, $\phi_q^n(b) = b$ for any $b \in \mathbb{F}_{q^n}$. So $\phi_q^n = \mathrm{id}$. The rest is clear. □

[5] L.X. thanks Professor Jiangwei Xue for pointing out an error in an earlier version of this note. Indeed, $\mathrm{Gal}(K/\mathbb{Q})$ could be a strict subgroup of $\mathbf{Z}_n \rtimes \mathbb{Z}_n^\times$. For example, consider the case $K = \mathbb{Q}(\sqrt[10]{5}, \zeta_{10})$; it is Galois extension of degree 20 over $\mathbb{Q}$ whereas $\#(\mathbb{Z}_{10} \rtimes \mathbb{Z}_{10}^\times) = 40$. This is because $\mathbb{Q}(\sqrt{5})$ is contained as a subfield of $\mathbb{Q}(\zeta_{10})$; so if an automorphism of K fixes $\sqrt[10]{5}$, it also fixes $\sqrt{5}$ and can only sends ζ_5 to $\zeta_5^{\pm 1}$. So the Galois group $\mathrm{Gal}(K/\mathbb{Q})$ is a subgroup of $\mathbf{Z}_{10} \rtimes \mathbf{Z}_{10}^\times$.

Example 17.3.3. We give an example of the Galois diagram for extensions of finite fields.

$$\begin{tikzcd}
 & \mathbb{F}_{q^{12}} \arrow[dl, dash] \arrow[dr, dash] & & \\
\mathbb{F}_{q^4} \arrow[dr, dash] & & \mathbb{F}_{q^6} \arrow[dl, dash] \arrow[dr, dash] & \\
 & \mathbb{F}_{q^2} \arrow[dr, dash] & & \mathbb{F}_{q^3} \arrow[dl, dash] \\
 & & \mathbb{F}_q &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \{1\} \arrow[dl, dash] \arrow[dr, dash] & & \\
\langle \phi_q^4 \rangle \arrow[dr, dash] & & \langle \phi_q^6 \rangle \arrow[dl, dash] \arrow[dr, dash] & \\
 & \langle \phi_q^2 \rangle \arrow[dr, dash] & & \langle \phi_q^3 \rangle \arrow[dl, dash] \\
 & & \langle \phi_q \rangle \cong \mathbf{Z}_{12} &
\end{tikzcd}$$

$$\phi_q \longleftrightarrow 1$$

17.4. **Cyclotomic extension.**

Definition 17.4.1. For a positive integer $n \in \mathbb{N}$,

$$\mu_n := \big\{ n\text{th roots of unity in } \mathbb{C} \big\} = \langle \zeta_n \rangle \cong \mathbf{Z}_n,$$

where $\zeta_n = e^{2\pi \mathbf{i}/n}$.

Define $\mathbb{Q}(\mu_n) = \mathbb{Q}(\zeta_n) \subseteq \mathbb{C}$; it is a finite field extension of $\mathbb{Q}$, called the n**th cyclotomic extension of** $\mathbb{Q}$.

A **primitive** n**th root of unity** is a generator of μ_n; it is equal to ζ_n^a for some $a \in \mathbf{Z}_n^\times$. Define

$$\Phi_n(x) := \prod_{a \in \mathbf{Z}_n^\times} \big(x - \zeta_n^a \big);$$

it is called the n**th cyclotomic polynomial**.

Example 17.4.2. We have $\Phi_1(x) = x - 1$, $\Phi_2(x) = x + 1$, $\Phi_3(x) = x^2 + x + 1$.

Lemma 17.4.3. *We have*

$$x^n - 1 = \prod_{d|n} \Phi_d(x).$$

Each $\Phi_n(x)$ is a polynomial of degree $\varphi(n)$ with coefficients in $\mathbb{Z}$.

Proof. The first equality is easy:

$$x^n - 1 = \prod_{b \in \mathbf{Z}_n} \big(x - \zeta_n^b \big) = \prod_{d|n} \prod_{i \in \mathbf{Z}_d^\times} \big(x - \zeta_n^{di} \big) = \prod_{d|n} \Phi_d(x). \tag{17.4.3.1}$$

We will prove that $\Phi_n(x)$ has coefficients in $\mathbb{Z}$ and its coefficients have gcd $= 1$. Assume that this has been proved for smaller n. Then (17.4.3.1) and Gauss' lemma implies that $\Phi_n(x)$ has coefficients in $\mathbb{Z}$ and has coefficients' gcd $= 1$. □

Theorem 17.4.4. *The polynomial $\Phi_n(x)$ is irreducible in $\mathbb{Q}[x]$. So $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$.*

Proof. It suffices to show that $\Phi_n(x)$ is irreducible in $\mathbb{Z}[x]$. Let ζ be a primitive nth root of unity in a splitting field of $\Phi_n(x)$. (We deliberately do not specify one here.)

We need to show that the minimal polynomial $f(x) := m_{\zeta,\mathbb{Q}}(x)$ of ζ over $\mathbb{Q}$ is equal to $\Phi_n(x)$; it is clear that $f(x)|\Phi_n(x)$. We will show that for any integer a relatively prime to n, ζ^a is a zero of $\Phi_n(x)$.

We take a prime p *not* dividing n.

Claim: ζ^p is also a zero of $f(x)$.

This claim would imply that if $a = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ is relatively prime to n, $\zeta^a = \zeta^{p_1^{\alpha_1} \cdots p_r^{\alpha_r}}$. Iteratively, we can prove that ζ is a zero of $f(x)$ implying ζ^a is a zero of $f(x)$. From this, we deduce that $f(x) = \Phi_n(x)$. (This is why we did not specify a primitive nth root of unity.)

Now, we prove the Claim. Suppose this is not true. Let $g(x) = m_{\zeta^p, \mathbb{Q}}(x)$ be the minimal polynomial of ζ^p over $\mathbb{Q}$.

As $f(x) \neq g(x)$, we have $\gcd(f(x), g(x)) = 1$ and thus

$$f(x)g(x) \mid \Phi_n(x).$$

On the other hand, $g(\zeta^p) = 0$ implies that ζ is a zero of $g(x^p)$. This implies that

$$f(x) \mid g(x^p) \quad \Rightarrow \quad g(x^p) = f(x)h(x) \text{ in } \mathbb{Z}[x],$$

for some $h(x) \in \mathbb{Z}[x]$. Taking this equation modulo p, we have

$$\bar{g}(x)^p = \bar{g}(x^p) = \bar{f}(x)\bar{h}(x) \quad \text{in } \mathbb{F}_p[x].$$

This implies that $\bar{f}(x)$ and $\bar{g}(x)$ have a common factor in $\mathbb{F}_p[x]$.

Yet $\bar{f}(x)\bar{g}(x)$ divides $\bar{\Phi}_n(x)$, which further divides $x^n - 1$. This implies that $x^n - 1$ has repeated zeros in its splitting field over $\mathbb{F}_p$. But

$$\big(x^n - 1, D(x^n - 1)\big) = \big(x^n - 1, nx^{n-1}\big) = \big(x^n - 1, x^{n-1}\big) = (1).$$

This contradicts with the properties of repeated zeros. The Claim is proved.

This completes the proof of irreducibility of $\Phi_n(x)$. □

The following is clear.

Corollary 17.4.5. *The Galois group of* $\mathbb{Q}(\zeta_n)/\mathbb{Q}$ *is* $\mathbf{Z}_n^\times$. *Explicitly, for* $a \in \mathbf{Z}_n^\times$*, the associated automorphism is*

$$\begin{tikzcd}
\mathbb{Q}(\zeta_n) & \mathbb{Q}[x]/(\Phi_n(x)) \arrow[l, "\cong"'] \arrow[r, "\cong"] & \mathbb{Q}(\zeta_n) \\
\zeta_n & x + \Phi_n(x), \arrow[l, mapsto] \arrow[r, mapsto] & \zeta_n^a.
\end{tikzcd}$$

Corollary 17.4.6. *For every finite abelian group G, there exists a finite Galois extension K of $\mathbb{Q}$ with Galois group G.*

Proof. Write $G = \mathbf{Z}_{n_1} \times \cdots \times \mathbf{Z}_{n_r}$. For each n_i, find a distinct odd prime number p_i such that $p_i \equiv 1 \bmod n_i$. Then G is a quotient of

$$\mathbf{Z}_{p_1}^\times \times \cdots \times \mathbf{Z}_{p_r}^\times \simeq \mathbf{Z}_{p_1 - 1} \times \cdots \times \mathbf{Z}_{p_r - 1}.$$

Say the kernel of this quotient is H. Then we consider the field extensions:

$$\begin{tikzcd}
\mathbb{Q}(\zeta_{p_1 \cdots p_r}) \arrow[dd, dash] \arrow[dd, dash, bend right, "\prod_i \mathbb{Z}_{p_i}^\times"'] \arrow[dr, dash, "H"] & \\
& K = \mathbb{Q}(\zeta_{p_1 \cdots p_r})^H \arrow[dl, dash, "G"] \\
\mathbb{Q} &
\end{tikzcd}$$

The field $K = \mathbb{Q}(\zeta_{p_1 \cdots p_r})^H$ is what we seek for. □

Example 17.4.7. We illustrate the above proof by constructing a cyclic extension of $\mathbb{Q}$ or degree 3. Write $\zeta = \zeta_7$. Then we have

$$\begin{tikzcd}
\mathbf{Z}_7^\times \arrow[r, phantom, "\cong"] & \mathbf{Z}_6 \arrow[r, "\phi"] & \mathbf{Z}_3 \\
\{1,-1\} \arrow[u, phantom, "\bigcup"] \arrow[r, leftrightarrow] & \ker\phi = \{0,3\} \arrow[u, phantom, "\bigcup"] &
\end{tikzcd}$$

But we have to translate $\ker\phi$ in terms of $\mathbf{Z}_7^\times$, it is the subset $\{1,-1\}$ (namely $\{x \in \mathbf{Z}_7^\times \mid x^2 = 1 \bmod 7\}$). Taking trace, we have

$$\theta := \zeta + \zeta^{-1} \in \mathbb{Q}(\zeta)^{\{1,-1\}}.$$

We may compute the minimal polynomial of θ as follows:

$$\begin{aligned}
\theta^2 &= \zeta^2 + \zeta^{-2} + 2. \\
\theta^3 &= \zeta^3 + \zeta^{-3} + 3(\zeta + \zeta^{-1}).
\end{aligned}$$

Using the fact that $1 + \zeta + \zeta^2 + \zeta^3 + \zeta^{-3} + \zeta^{-2} + \zeta^{-1} = 0$, we deduce that

$$\theta^3 + \theta^2 = \zeta^3 + \zeta^{-3} + \zeta^2 + \zeta^{-2} + 3(\zeta + \zeta^{-1}) + 2 = 2(\zeta + \zeta^{-1}) + 1 = 2\theta + 1.$$

So the minimal polynomial of θ is $x^3 + x^2 - 2x - 1$.

The following is a converse of Corollary 17.4.6. It is one of the first achievement in the history of number theory, marking a starting point of the study fo abelian extensions of number fields.

Theorem 17.4.8 (Kronecker–Weber)**.** *Every finite abelian extension K of $\mathbb{Q}$ is contained in some $\mathbb{Q}(\zeta_n)$.*

Extended reading after Lecture 17

In Example 17.2.4, we in fact used a standard tool to produce elements in a fixed field, called the "trace" map. We give a formal definition here.

Definition 17.4.9. For a finite extension K of F of degree n and $a \in K$, we define its trace and norm over F as follows: viewing K as a finite dimensional F-vector space (and choose a basis), then multiplication by a is a F-linear map T_a on K given by an $n \times n$ matrix (with coefficients in F). Then the **trace** and the **norm** of a over F is

$$\mathrm{Tr}(a) = \mathrm{Tr}_{K/F}(a) = \mathrm{Tr}(T_a) \quad \text{and} \quad \mathrm{Nm}(a) = \mathrm{Nm}_{K/F}(a) = \det(T_a).$$

As the traces and determinants of matrices do not depend on the choice of the basis, the traces and the norms are well-defined.

Clearly, for $\alpha, \beta \in K$, we have

$$\mathrm{Tr}_{K/F}(\alpha + \beta) = \mathrm{Tr}_{K/F}(\alpha) + \mathrm{Tr}_{K/F}(\beta) \quad \text{and} \quad \mathrm{Nm}_{K/F}(\alpha\beta) = \mathrm{Nm}_{K/F}(\alpha)\mathrm{Nm}_{K/F}(\beta).$$

Lemma 17.4.10. *If K is a finite extension of F and E an intermediate field, we have for $\alpha \in K$,*

$$\mathrm{Tr}_{K/F}(\alpha) = \mathrm{Tr}_{E/F}(\mathrm{Tr}_{K/E}(\alpha)) \quad \textit{and} \quad \mathrm{Nm}_{K/F}(\alpha) = \mathrm{Nm}_{E/F}(\mathrm{Nm}_{K/E}(\alpha)).$$

Proof. We leave this as an exercise. □

Lemma 17.4.11. *Let K be a finite extension over F of degree n, and let $\alpha \in K$ be an element with minimal polynomial $m_{\alpha,F}(x) = x^m + a_1x^{m-1} + \cdots + a_m \in F[x]$ with $m|n$. Then*

$$\mathrm{Tr}_{K/F}(\alpha) = -\frac{n}{m}a_1 \quad \textit{and} \quad \mathrm{Nm}_{K/F}(\alpha) = (-1)^n(a_m)^{n/m}.$$

Moreover, when K is Galois over F with Galois group G, we have

$$\mathrm{Tr}_{K/F}(\alpha) = \sum_{g\in G} g(\alpha) \quad \textit{and} \quad \mathrm{Nm}_{K/F}(\alpha) = \prod_{g\in G} g(\alpha).$$

Proof. We first treat the case when $K = F(\alpha)$ (in this case $m = n$). In this case, $K \cong F[x]/(m_{\alpha,F}(x))$ with basis elements $1, x, \dots, x^{n-1}$, and multiplication by α is represented by the matrix

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 1 & 0 & \ldots & 0 & -a_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -a_1 \end{pmatrix}$$

It is clear that $\mathrm{Tr}_{K/F}(\alpha) = -a_1$ and $\det_{K/F}(\alpha) = (-1)^n a_n$. In the case when K/F is Galois, $\{g(\alpha) \,|\, g \in G\}$ are all zeros of $m_{\alpha,F}(x)$. Thus, we have

$$\mathrm{Tr}_{K/F}(\alpha) = \sum_{g\in G} g(\alpha) \quad \text{and} \quad \mathrm{Nm}_{K/F}(\alpha) = \prod_{g\in G} g(\alpha).$$

In general, consider the tower of extensions

$$\begin{tikzcd} K \arrow[d, dash] \\ F(\alpha) \arrow[d, dash] \\ F \end{tikzcd}$$

We deduce immediately from Lemma 17.4.10 that

$$\begin{aligned} \mathrm{Tr}_{K/F}(\alpha) &= \mathrm{Tr}_{F(\alpha)/F}\big(\mathrm{Tr}_{K/F(\alpha)}(\alpha)\big) = \mathrm{Tr}_{F(\alpha)/F}\Big(\frac{n}{m}\alpha\Big) = -\frac{n}{m}a_1, \\ \mathrm{Nm}_{K/F}(\alpha) &= \mathrm{Nm}_{F(\alpha)/F}\big(\mathrm{Nm}_{K/F(\alpha)}(\alpha)\big) = \mathrm{Nm}_{F(\alpha)/F}\big(\alpha^{n/m}\big) = (-1)^n a_m^{n/m}. \end{aligned}$$

When K is a Galois extension over F with Galois group G, through Galois theory, $F(\alpha) = K^H$ for some subgroup $H \leq G$, we have

$$\sum_{g\in G} g(\alpha) = \frac{n}{m} \sum_{gH\in G/H} g(\alpha) = \frac{n}{m}\cdot(-a_1) \quad \text{and} \quad \prod_{g\in G} g(\alpha) = \Big(\sum_{gH\in G/H} g(\alpha)\Big)^{n/m} = (-1)^n a_m^{n/m}.$$

The lemma is proved. □

18. Galois theory II: proof

The goal of this lecture is to prove the main theorem of Galois theory.

18.1. **Proof of Galois theory.**

Before giving the proof, we quickly recall an important result we proved earlier.

Proposition 18.1.1. *Let K be a finite Galois (namely normal and separable) extension of F, and let L be an extension of K normal over F.*

$$\begin{tikzcd} K \arrow[rr, dashed] \arrow[dr, dash] & & L \arrow[dl, dash] \\ & F & \end{tikzcd}$$

We have the following equalities

$$\begin{array}{ccccccc} [K:F] & = & \#\mathrm{Hom}_F(K,L) & = & \#\mathrm{Hom}_F(K,K) & = & \#\mathrm{Gal}(K/F). \\ & \uparrow & & \uparrow & & & \\ & \textit{K/F separable} & & \begin{array}{c}\textit{K/F normal}\\ \textit{so can take } K=L\end{array} & & & \end{array}$$

We now give the proof of the main theorem of Galois theory.

Theorem 18.1.2 (Galois theory)**.** *Let K be a finite Galois extension with $G = \mathrm{Gal}(K/F)$.*

(1) *Then there is a one-to-one correspondence between*

$$\begin{array}{ccc} \{\textit{subgroups } H \le G\} & \longleftrightarrow & \{\textit{intermediate fields } E \textit{ of } K/F\} \\ H & \longmapsto & K^H = \{x \in K \mid h(x) = x \textit{ for any } h \in H\} \\ \mathrm{Gal}(K/E) = \{g \in G \mid g|_E = \mathrm{id}|_E\} & \longleftarrow\!\dashv & E \end{array}$$

Proof. (1) Since K is a finite normal extension of F, K is a splitting field for some $f(x) \in F[x]$. (This implies that K is also the splitting field of $f(x)$ over an intermediate field E.) It follows from Proposition 18.1.1 that we have

$$[K:E] = \mathrm{Gal}(K/E).$$

(a) Given a subgroup $H \le G$, we need to show that $\mathrm{Gal}(K/K^H) = H$.

Clearly, for any $h \in H$, h fixes K^H and thus $H \subseteq \mathrm{Gal}(K/K^H)$. It remains to show that

$$\#H \ge \#\mathrm{Gal}(K/K^H) = [K : K^H]. \tag{18.1.2.1}$$

Now we use the primitive element theorem (Theorem 16.3.1) to the separable extension K/K^H to see that $K = K^H(\alpha)$ for some $\alpha \in K$. Then for the minimal polynomial $m_{\alpha,K^H}(x) \in K^H[x]$ of α over K^H, we have $[K : K^H] = \deg m_{\alpha,K^H}(x)$.

Yet we may consider another polynomial

$$f(x) = \prod_{h\in H}(x - h(\alpha)) \in K[x].$$

The coefficients can be seen to actually belong to K^H. By basic properties of minimal polynomial, $m_{\alpha,K^H}(x)$ divides $f(x)$ and in particular,

$$[K : K^H] = \deg m_{\alpha,K^H}(x) \le \deg f(x) = \#H.$$

This completes the proof of $\mathrm{Gal}(K/K^H) = H$.

(b) Given intermediate field E of K over F, we need to show that $K^{\mathrm{Gal}(K/E)} = E$. First of all, $E \subseteq K^{\mathrm{Gal}(K/E)}$ because any $h \in \mathrm{Gal}(K/E)$ fixes E. Next, we count

$$[K:E] \underset{\substack{\uparrow \\ \text{Proposition 18.1.1}}}{=} \#\mathrm{Gal}(K/E) \underset{\substack{\uparrow \\ \text{proved in (a)}}}{=} [K:K^{\mathrm{Gal}(K/E)}].$$

This implies that $E = K^{\mathrm{Gal}(K/E)}$. □

Remark 18.1.3. In fact, if we inspect the proof of Galois theory, the only non-formal argument is in (1)(a), we need to show that $\#H \geq [K:K^H]$, namely (18.1.2.1). There are typically two proofs for this; we presented the one using primitive element theorem. There is another proof of this using a lemma of Artin, which we present here in the following lemma (part (1)).

Lemma 18.1.4. *Assume that a finite group G acts on a field K via automorphisms. Put $F = K^G$. Then*

(1) *$\#G \geq [K:F]$.*

(2) *K is a Galois extension of F with Galois group G. In particular, $\#G = [K:F]$.*

Proof. (1) Let $n = \#G$ and write $G = \{\sigma_1 = \mathrm{id}, \sigma_2, \dots, \sigma_n\}$. To prove that $[K:F] \leq n$, we need to show that every $n+1$ elements $u_1, \dots, u_{n+1}$ of K are F-linearly dependent. Write

$$A := \begin{pmatrix} \sigma_1(u_1) & \cdots & \sigma_1(u_{n+1}) \\ \vdots & \ddots & \vdots \\ \sigma_n(u_1) & \cdots & \sigma_n(u_{n+1}) \end{pmatrix} \in \mathrm{Mat}_{n\times(n+1)}(K).$$

Automatically, the column vectors $\vec{v}_1, \dots, \vec{v}_{n+1}$ are K-linearly dependent. So there exists some r such that $\vec{v}_1, \dots, \vec{v}_r$ are K-linearly independent and $\vec{v}_1, \dots, \vec{v}_{r+1}$ are not. Then we must have

$$\vec{v}_{r+1} = \alpha_1 \vec{v}_1 + \cdots + \alpha_r \vec{v}_r. \tag{18.1.4.1}$$

We hope to show that each α_i in fact belongs to F (or equivalently, $\sigma(\alpha_i) = \alpha_i$ for every $\sigma \in G$), then $\vec{v}_i$'s are F-linearly dependent, which implies that u_i's are F-linearly dependent, proving our result.

We apply $\sigma \in \mathrm{Aut}(K/F)$ to (18.1.4.1) to get

$$\sigma(\vec{v}_{r+1}) = \sigma(\alpha_1)\sigma(\vec{v}_1) + \cdots + \sigma(\alpha_r)\sigma(\vec{v}_r).$$

But the column vector $\sigma(\vec{v}_i)$ is just

$$\sigma \begin{pmatrix} \sigma_1(u_i) \\ \vdots \\ \sigma_n(u_i) \end{pmatrix} = \begin{pmatrix} \sigma\sigma_1(u_i) \\ \vdots \\ \sigma\sigma_n(u_i), \end{pmatrix}$$

which simply permutes the entries. This then implies that

$$\vec{v}_{r+1} = \sigma(\alpha_1)\vec{v}_1 + \cdots + \sigma(\alpha_r)\vec{v}_r. \tag{18.1.4.2}$$

But (18.1.4.1) and (18.1.4.2) must be the same relation!! This implies that

$$\text{for each } i, \quad \sigma(\alpha_i) = \alpha_i.$$

As this works for all $\sigma \in G$, we deduce $\alpha_i \in F$, proving part (1).

(2) We show that K is a separable and normal extension of F. Let $f(x) \in F[x]$ be an irreducible polynomial that has a zero α in K. Write $H = \{g \in G \mid g(\alpha) = \alpha\}$. Then

$$h(x) := \prod_{gH \in G/H} (x - g(\alpha)).$$

(Since $H\alpha = \alpha$, this product makes sense.) The coefficients of $h(x)$ are symmetric polynomials in $g(\alpha)$ for all $gH \in G/H$; so they are invariant under the left action by any element $\sigma \in G$. So $h(x) \in K^G[x] = F[x]$.

Moreover, since $(h(x), f(x)) \neq (1)$ when considered in $K[x]$ (as they have a common zero α), we have $(h(x), f(x)) \neq (1)$ in $F[x]$. As $f(x)$ is irreducible, $f(x) \mid h(x)$. In particular, $f(x)$ splits completely in $K[x]$ and has only simple zeros, because $h(x)$ has.

Thus, K is separable and normal over F, i.e. is Galois over F. In particular, $G \subseteq \mathrm{Gal}(K/F)$, and thus

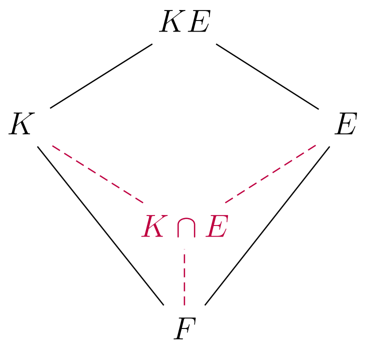
$$\#G \leq \#\mathrm{Gal}(K/F) \overset{\text{Lemma 18.1.1}}{=} [K : F].$$

Combining part (1) shows that $\#G = [K : F]$.

□

18.2. **Galois theory for composite of fields.** We are interested in how Galois theory behave under composite of fields.

Proposition 18.2.1. *Consider the following diagram of field extensions.*



Assume that K is a finite Galois extension and E is an arbitrary field extension of F (not even assume to be algebraic). Then KE is a Galois extension over E and

$$\mathrm{Gal}(KE/E) \cong \mathrm{Gal}(K/K \cap E).$$

As a corollary, if E is a finite extension over F, then we have

$$[KE : K \cap E] = [K : K \cap E] \cdot [E : K \cap E].$$

Caveat 18.2.2. We have seen earlier that this proposition would be false without the assumption that K is Galois over F. One only has an inequality

$$[KE : K \cap E] \geq [K : K \cap E] \cdot [E : K \cap E].$$

A typical example is the following:

$$\begin{tikzcd}
& \mathbb{Q}(\sqrt[3]{2}, \zeta_3) \arrow[dl, dash] \arrow[dr, dash] & \\
\mathbb{Q}(\sqrt[3]{2}) \arrow[dr, dash] & & \mathbb{Q}(\sqrt[3]{2}\zeta_3) \arrow[dl, dash] \\
& \mathbb{Q} &
\end{tikzcd}$$

So what happened? Suppose that $L := KE$ is a Galois extension of F with Galois group $G = \mathrm{Gal}(L/F)$ and that $F = K \cap E$. We have the following diagram

$$\begin{tikzcd}
& L := KE \arrow[dl, dash] \arrow[dr, dash] \arrow[dl, dashed, dash, bend right, "H_2"'] \arrow[dr, dashed, dash, bend left, "H_1"] & \\
K \arrow[dr, dash] & & E \arrow[dl, dash] \\
& F &
\end{tikzcd}$$

Set $H_1 = \mathrm{Gal}(L/K)$ and $H_2 = \mathrm{Gal}(L/E)$. Thus

$$F = K \cap E = L^{H_1} \cap L^{H_2} = L^{\langle H_1, H_2 \rangle} \ \Rightarrow\ \langle H_1, H_2 \rangle = G.$$

On the other hand, we have

$$KE = L \ \Rightarrow\ H_1 \cap H_2 = \{1\}.$$

The key here is that $H_1 H_2 \subseteq \langle H_1, H_2 \rangle = G$ and the first inclusion is typically strict, and thus

$$\#G \geq \#H_1 \cdot \#H_2 \ \Rightarrow\ [L : F] \geq [L : K] \cdot [L : E] \ \Rightarrow\ [E : F] \cdot [K : F] \geq [L : F].$$

However, if one of H_i is a normal subgroup of G, $\langle H_1, H_2 \rangle = H_1 H_2 = G$. The equalities above hold.

Proof of Proposition 18.2.1. Since K is a finite Galois extension of F, K is the splitting field of some separable polynomial $f(x)$ over F. This further implies that KE is the splitting field of the same polynomial $f(x)$ over E. This implies that KE is a Galois extension of E.

Moreover, there is a natural homomorphism

$$\begin{aligned} \Psi : \mathrm{Gal}(KE/E) &\longrightarrow \mathrm{Gal}(K/K \cap E). \\ \sigma &\longmapsto \sigma|_K \end{aligned}$$

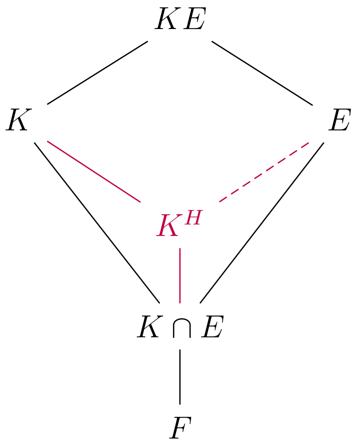
Here we used that K is normal over F; so it is stable under the action of any element of $\sigma \in \mathrm{Gal}(KE/E)$. We aim to show that Ψ is an isomorphism.

We compute the kernel of Ψ:

$$\ker(\Psi) = \big\{ \sigma \in \mathrm{Gal}(KE/E) \,\big|\, \sigma|_K = \mathrm{id} \big\}.$$

But for such σ, $\sigma|_E = \mathrm{id}$ and $\sigma|_K = \mathrm{id}$, so σ is trivial when restricted to KE and thus $\sigma = 1$. This shows that Ψ is injective.

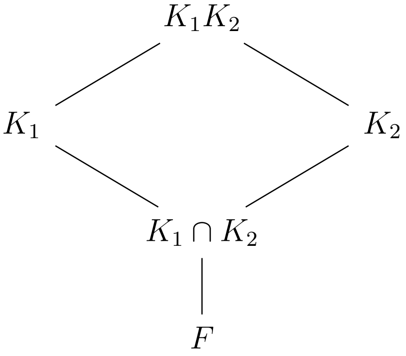
Now we prove the surjectivity of Ψ. Let $H := \mathrm{Im}(\Psi) \subseteq \mathrm{Gal}(K/K \cap E)$ be the subgroup. Consider $K^H \supseteq K \cap E$.



If we can show that $K^H \subseteq E$, then $K^H \subseteq K \cap E$, and we are forced to have an equality, and thus Ψ is surjective.

We note that, for any $\sigma \in \mathrm{Gal}(KE/E)$, we have $\sigma|_{K^H} = \mathrm{id}$ and $\sigma|_E = \mathrm{id}$ and thus $\sigma|_{K^H E} = \mathrm{id}$. Thus $K^H E$ is fixed by $\mathrm{Gal}(KE/E)$. This means that $K^H E = E$, which means $K^H \subseteq E$. The proposition is proved. □

Proposition 18.2.3. *Suppose that we have a tower of extensions below, in which K_1 and K_2 are Galois over F.*



Then we have

(1) *$K_1 \cap K_2$ is Galois over F.*

(2) *K_1K_2 is Galois over F and*

$$\mathrm{Gal}(K_1K_2/F) = \big\{(g_1, g_2) \in \mathrm{Gal}(K_1/F) \times \mathrm{Gal}(K_2/F) \,\big|\, g_1|_{K_1 \cap K_2} = g_2|_{K_1 \cap K_2}\big\}.$$

(In particular, if $K_1 \cap K_2 = F$, we have $\mathrm{Gal}(K_1K_2/F) = \mathrm{Gal}(K_1/F) \times \mathrm{Gal}(K_2/F)$.*)*

Proof. (1) We need to show that $K_1 \cap K_2$ is normal over F. Suppose that $f(x) \in F[x]$ is a polynomial that has a zero in $K_1 \cap K_2$. Then by normality of $f(x)$, all zeros of $f(x)$ belong to both K_1 and K_2 and thus $f(x)$ splits over $K_1 \cap K_2$. So $K_1 \cap K_2$ is normal (and hence Galois) over F.

(2) Let K_i be the splitting field of a separable polynomial $f_i(x)$ for $i = 1, 2$. This means that K_1K_2 is a splitting field of $f_1(x)f_2(x)$. So K_1K_2 is Galois over F.

Consider the homomorphism

$$\begin{aligned} \varphi : \mathrm{Gal}(K_1K_2/F) &\longrightarrow \mathrm{Gal}(K_1/F) \times \mathrm{Gal}(K_2/F). \\ \sigma &\longmapsto (\sigma|_{K_1}, \sigma|_{K_2}) \end{aligned}$$

(Note that σ stabilizes each K_i because K_i is normal over F.)

We compute the kernel and the image of φ as follows:

$$\begin{aligned} \ker \varphi &= \big\{\sigma \in \mathrm{Gal}(K_1K_2/F) \,\big|\, \sigma|_{K_1} = \mathrm{id},\, \sigma|_{K_2} = \mathrm{id}\big\} = \{\mathrm{id}\}, \\ \mathrm{Im}(\varphi) &\subseteq \big\{(\sigma_1, \sigma_2) \in \mathrm{Gal}(K_1/F) \times \mathrm{Gal}(K_2/F) \,\big|\, \sigma_1|_{K_1\cap K_2} = \sigma_2|_{K_1\cap K_2}\big\}. \end{aligned}$$

Write A for the latter group, which contains $\mathrm{Gal}(K_1K_2/F)$ as as subgroup. Now we count.

$$\begin{aligned} \#\mathrm{Gal}(K_1K_2/F) &= [K_1K_2 : F] = [K_1K_2 : K_2] \cdot [K_2 : F] \\ &= [K_1 : K_1 \cap K_2] \cdot [K_2 : F] \qquad \text{(by Proposition 18.2.1)} \\ &= [K_1 : K_1 \cap K_2] \cdot [K_2 : K_1 \cap K_2] \cdot [K_1 \cap K_2 : F] \\ &= \#\mathrm{Gal}(K_1/K_1 \cap K_2) \cdot \#\mathrm{Gal}(K_2/K_1 \cap K_2) \cdot \#\mathrm{Gal}(K_1 \cap K_2/F) \\ &= \#A. \end{aligned}$$

The proposition is proved. □

Example 18.2.4. As an application, we see that $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7})$ is a Galois extension of $\mathbb{Q}$ with Galois group $\mathbf{Z}_2^4$.

18.3. **Linear independence of characters.** We end this section with a discussion on Artin's linear independence of characters theorem. This theorem has very interesting applications in number theory, and we will encounter one of them in the next lecture.

Definition 18.3.1. Let H be an abelian group and let L be a field. A **character** χ of H with values in L is a group homomorphism

$$\chi : H \to L^\times$$

This is the same as "1-dimensional representations" of H with coefficients in L.

Definition 18.3.2. We say that characters $\chi_1, \dots, \chi_n$ are **linearly independent** over L if they are linearly independent as functions on H, i.e. for any $a_1, \dots, a_n \in L$, if $a_1\chi_1(h) + \cdots + a_n\chi_n(h) = 0$ for any $h \in H$, then $a_1 = \cdots = a_n = 0$.

Theorem 18.3.3 (Linearly independence of characters)**.** *If $\chi_1, \dots, \chi_n$ are distinct characters of a group H with values in L, then they are linearly independent.*

Proof. Suppose that these characters $\chi_1, \dots, \chi_n$ are linearly dependent. Then among all (nonzero) linear relations, there is a unique one with minimal number of nonzero a_i's. Without loss of generality, we assume this is

$$a_1\chi_1 + a_2\chi_2 + \cdots + a_r\chi_r = 0 \quad \text{as functions on } H.$$

In other words, for any $h \in H$,

$$a_1\chi_1(h) + \cdots + a_r\chi_r(h) = 0. \tag{18.3.3.1}$$

Since $\chi_1 \neq \chi_r$, there exists $h_0 \in H$ such that $\chi_1(h_0) \neq \chi_r(h_0)$. Applying (18.3.3.1) to hh_0, we have

$$\begin{aligned} a_1\chi_1(h_0h) + \cdots + a_r\chi_r(h_0h) &= 0, \\ a_1\chi_1(h_0)\chi_1(h) + \cdots + a_r\chi_r(h_0)\chi_r(h) &= 0. \end{aligned}$$

Taking the difference of this and $\chi_1(h_0)$ times (18.3.3.1), we deduce

$$\underbrace{a_2(\chi_1(h_0) - \chi_2(h_0))}_{b_2}\chi_2(h) + \cdots + \underbrace{a_r(\chi_1(h_0) - \chi_r(h_0))}_{b_r}\chi_r(h) = 0.$$

This gives another nontrivial linear relation among χ_i's (noting that $b_r \neq 0$) with smaller number of nonzero coefficients. □

18.3.4. *Application of Theorem 18.3.3 to Galois theory.* We will apply the linear independence result in the following way: let K be a finite extension of F and let L be another extension of F. Suppose that $\sigma_1, \dots, \sigma_n \in \mathrm{Hom}_F(K, L)$ be distinct embeddings.

$$\begin{tikzcd} K \arrow[rr, hook, "\sigma_i"] \arrow[dr, dash] & & L \arrow[dl, dash] \\ & F & \end{tikzcd}$$

If we write $K = Fe_1 \oplus Fe_2 \oplus \cdots \oplus Fe_r$ for $r = [K : F]$, we consider the following matrix

$$A = \begin{pmatrix} \sigma_1(e_1) & \cdots & \sigma_1(e_r) \\ \vdots & \ddots & \vdots \\ \sigma_n(e_1) & \cdots & \sigma_n(e_r) \end{pmatrix} \in \mathrm{Mat}_{n\times r}(L).$$

The linearly independence theorem implies that the rows of this matrix A is L-linearly independent. This implies that $r \geq n$.

Remark 18.3.5. This gives another proof of $[K : F] \geq \#\mathrm{Hom}_F(K, L)$.

The following is a side result that is useful in many applications.

Lemma 18.3.6. *Let K be a finite separable field extension of F so that $K = F[x]/(f(x))$ and let L be normal extension of F containing K. Then we have an isomorphism of L-algebras.*

$$\begin{aligned} L \otimes_F K := L[x]/(f(x)) &\xrightarrow[\cong]{\varphi = (\varphi_\sigma)_\sigma} \textstyle\prod_{\sigma \in \mathrm{Hom}_F(K,L)} L \\ a \otimes f(x) &\longmapsto \big(a\sigma(f(x))\big)_\sigma. \end{aligned}$$

(We did not get to properly define tensor product; so for this problem, we think of $K \otimes_F L$ as just $L[x]/(f(x))$.)

Example 18.3.7.

$$\begin{aligned} \mathbb{R} \otimes_{\mathbb{Q}} \mathbb{Q}(\sqrt{2}) \cong \mathbb{R}[x]/(x^2 - 2) &\xrightarrow{\cong} \mathbb{R} \times \mathbb{R} \\ a + b\sqrt{2} &\longmapsto (a + b\sqrt{2}, a - b\sqrt{2}). \end{aligned}$$

Proof of Lemma 18.3.6. The map φ is clearly a well-defined homomorphism and L-linear. Both sides are L-vector spaces of dimension $[K : F] = \#\mathrm{Hom}_F(K, L)$. It suffices to show injectivity and this is exactly the linearly independence of characters above (for characters of $K^\times$). □

19. Galois group of polynomials and insolvability of quintics

19.1. **Cyclic extensions and Kummer theory.**

Definition 19.1.1. The extension K is called **cyclic** if K is a Galois extension over F and $\mathrm{Gal}(K/F)$ is cyclic.

Proposition 19.1.2. *Assume that*

(1) $\mathrm{char}(F)$ *does not divide* n*, and*
(2) F *contains all* n*th roots of unity.*

Then $K = F(\sqrt[n]{a})$ *is a cyclic extension of degree dividing* n.

Proof. We may factor

$$x^n - a = (x - \sqrt[n]{a})(x - \zeta_n \sqrt[n]{a}) \cdots (x - \zeta_n^{n-1} \sqrt[n]{a})$$

So K is the splitting field of $x^n - a$ over F. As $(x^n-a, D(x^n-a)) = (x^n-a, nx^{n-1}) = (a) = (1)$, $x^n - a$ is separable.

$$\begin{tikzcd} K = F(\sqrt[n]{a}) \arrow[d, dash] \\ F \end{tikzcd}$$

For each $\sigma \in \mathrm{Gal}(K/F)$, $\sigma(\sqrt[n]{a}) = \zeta_n^{\lambda(\sigma)} \sqrt[n]{a}$ for some $\lambda(\sigma) \in \mathbf{Z}_n$. This defines an *injective* map

$$\begin{aligned} \lambda : \mathrm{Gal}(K/F) &\longrightarrow \mathbf{Z}_n = \mu_n \\ \sigma &\longmapsto \lambda(\sigma). \end{aligned}$$

The injectivity follows from the fact that $\sqrt[n]{a}$ generates K over F.

This λ is a homomorphism because for $\tau, \sigma \in \mathrm{Gal}(K/F)$,

$$\zeta_n^{\lambda(\tau\sigma)} \sqrt[n]{a} = \tau\sigma(\sqrt[n]{a}) = \tau(\zeta_n^{\lambda(\sigma)} \sqrt[n]{a}) = \zeta_n^{\lambda(\sigma)} \zeta_n^{\lambda(\tau)} \sqrt[n]{a}.$$

So $\lambda(\tau\sigma) = \lambda(\tau) + \lambda(\sigma) \bmod n$; it is a homomorphism.

Through λ, $\mathrm{Gal}(K/F)$ may be viewed as a subgroup of $\mathbf{Z}_n$; in particular, $\mathrm{Gal}(K/F)$ is a cyclic group with order dividing n. □

In fact, when F contains nth roots of unity, all cyclic field extensions of F is of the form above. This is so-called **Kummer theory**.

Proposition 19.1.3 (Kummer). *Let* F *be a field such that* $\mathrm{char}(F) \nmid n$ *and assume that* F *contains all* n*th roots of unity. Then any cyclic field extension* K *of* F *of order* n *is of the form* $K = F(\sqrt[n]{a})$ *for some* $a \in F^\times$.

Proof. Write $\mathrm{Gal}(K/F) \cong \mathbf{Z}_n = \langle \sigma \rangle$. For each $\alpha \in K$, we define

$$b := \alpha + \zeta_n \sigma(\alpha) + \cdots + \zeta_n^{n-1} \sigma^{n-1}(\alpha). \tag{19.1.3.1}$$

(This is called the **Lagrange resolvent**.) By Theorem 18.3.3, $1, \sigma, \sigma^2, \dots, \sigma^{n-1}$ are linearly independent characters, there exists $\alpha \in K$ such that $b \neq 0$ (otherwise, the above expression gives a linear relations among $1, \sigma, \dots, \sigma^{n-1}$).

Note that, for such b, we have

$$\sigma(b) = \sigma(\alpha) + \zeta_n \sigma^2(\alpha) + \cdots + \zeta_n^{n-1} \sigma^n(\alpha) = \zeta_n^{-1} b.$$

(In fact, this is why we give such a definition to b.) So we have

$$\sigma(b^n) = (\zeta_n^{-1} b)^n = b^n.$$

Thus $a = b^n \in K$, and b can be viewed as an n-th root of a in K.

Moreover, for any σ^i, $\sigma^i(b) = \zeta_n^{-i} b$; so b is not contained in any intermediate fields between K and F. Thus $K = F(b) = F(\sqrt[n]{a})$. □

19.2. Expressing algebraic numbers by radicals. From now on, we assume that $\mathrm{char}(F) = 0$.

Definition 19.2.1. We say that an element α algebraic over F can be **expressed by radicals** or **solved in terms of radicals** over F if α belongs to some field K which is a succession of simple extensions

$$F = K_0 \subseteq K_1 \subseteq K_1 \subseteq \cdots \subseteq K_s = K, \tag{19.2.1.1}$$

such that $K_{i+1} = K_i(\sqrt[n_i]{a_i})$ for some $n_i \in \mathbb{N}$ and $a_i \in K_i$. The extensions K_{i+1}/K_i are called **radical extensions**.

Example 19.2.2. A typical such element α over $\mathbb{Q}$ is $\alpha = \sqrt[5]{\sqrt{5+\sqrt{7}} + \sqrt[4]{\sqrt{13}+\sqrt{17}}}$.

Proposition 19.2.3. *An element α can be expressed by radicals over a field F if α is contained in a* Galois *extension K of F which admits a tower of subfields of the form* (19.2.1.1).

Proof. By definition of expressing elements by radicals, there exists a finite extension K' of F which admits a tower of subfields of the form (19.2.1.1). Let K be the Galois closure of K' over F. This implies that for each $\sigma \in \mathrm{Hom}_F(K', L)$,

$$\begin{tikzcd}
K' = K'_s \arrow[d, dash] & & \sigma(K'_s) \arrow[d, dash] \\
\vdots \arrow[d, dash] & & \vdots \arrow[d, dash] \\
K'_1 \arrow[dr, dash] & & \sigma(K'_1) \arrow[dl, dash] \\
& F &
\end{tikzcd}$$

the composite $K'_s \cdot \sigma(K'_s)$ is an extension of F filtered by radical extensions. Continue this way proves the proposition. □

Theorem 19.2.4. *An (irreducible) polynomial $f(x)$ can be solved by radicals if and only if its Galois group (meaning the Galois group of its splitting field) is a solvable group.*

Proof. "⇒" As in the previous proposition, $f(x)$ splits over some Galois field L of F such that

$$F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r = L \quad \text{with} \quad K_{i+1} = K_i(\sqrt[n_i]{a_i}).$$

Define $F' := F(\zeta_{n_1}, \ldots, \zeta_{n_r})$, which is a Galois extension of F. Then we form extensions

$$\begin{tikzcd}
& F'L \arrow[d, dash] \\
K_r = L \arrow[ur, dash] \arrow[d, dash] & \vdots \arrow[d, dash] \\
\vdots \arrow[d, dash] & F'K_1 \arrow[d, dash] \\
K_1 \arrow[ur, dash] \arrow[d, dash] & F' \\
K_0 = F. \arrow[ur, dash] &
\end{tikzcd}$$

Then $F'L$ is Galois over F and each $F'K_{i+1} = F'K_i(\sqrt[n_i]{a_i})$ is Galois over $F'K_i$.

On the group side, Put $G = \mathrm{Gal}(F'L/F)$, $H_i = \mathrm{Gal}(F'L/F'K_i)$, and $H_0 = \mathrm{Gal}(F'L/F')$.

$$\begin{tikzcd}
& F'L \arrow[d, dash] \arrow[ddddl, dash, bend right=40, "G"'] \arrow[ddd, dash, bend right=50, "H_0"'] \arrow[dd, dash, bend left=50, "H_1"] \\
& \vdots \arrow[d, dash] \\
& F'K_1 \arrow[d, dash] \\
& F' \\
K_0 = F \arrow[ur, dash] &
\end{tikzcd}$$

As F' is a Galois extension of F, H_0 is a normal subgroup of G and G/H_0 is abelian. Similarly, $F'K_1/F'$ is Galois with cyclic Galois group by Proposition 19.1.2, then $H_1 \lhd H_0$ is a normal subgroup and H_0/H_1 is cyclic. We continue the discussion for each of $F'K_{i+1}/F'K_i$ to deduce that $H_{i+1} \lhd H_i$ and H_i/H_{i+1} is cyclic.

This implies that G is solvable, and hence $\mathrm{Gal}(L/F)$ as a quotient of G is solvable.

"$\Leftarrow$" Let K be a splitting field $f(x)$ over F. By Galois theory, we have a tower of intermediate fields

$$F = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_r = K$$

corresponding to subgroups $G = H_0 \supseteq H_1 \supseteq \cdots \supseteq H_r = \{1\}$ such that $H_{i+1} \rhd H_i$ and H_i/H_{i+1} is cyclic of order n_i. Again, put $F' = F(\zeta_{n_1}, \dots, \zeta_{n_r})$ and consider the composites

$$\begin{tikzcd}
 & F'K_r \\
K = K_r = L \arrow[ur, dash] \arrow[d, dash] & F'K_{r-1} \arrow[u, dash] \arrow[d, dash] \\
K_{r-1} \arrow[ur, dash] \arrow[d, dash] & \vdots \arrow[d, dash] \\
\vdots \arrow[d, dash] & F' \\
K_0 = F. \arrow[ur, dash] &
\end{tikzcd}$$

Then by Kummer theory, we have $F'K_{i+1} = F'K_i(\sqrt[n_i]{a_i})$ for some $n_i \in \mathbb{N}$ and $a_i \in K_i$, and thus α is solvable by radicals. □

Corollary 19.2.5. *If an equation has Galois group (in the sense of Definition 19.3.1 below) isomorphic to S_n or A_n with $n \geq 5$ (e.g. for a general irreducible polynomial of degree n), then it is not solvable by radicals.*

19.3. **Galois group of a polynomial.**

Definition 19.3.1. Let F be a field and $f(x) \in F[x]$ a separable polynomial. Let K be a splitting field of $f(x)$ over F. Then the **Galois group for** $f(x)$ is $\mathrm{Gal}(K/F)$.

Example 19.3.2. The Galois group for $x^7 - 5$ over $\mathbb{Q}$ (irreducible by Eisenstein criterion). The splitting field is $\mathbb{Q}(\sqrt[7]{5}, \zeta_7)$. The associated Galois group is $\mathbf{Z}_7 \rtimes \mathbf{Z}_7^\times$.

19.3.3. We ask a basic question: how to determine the Galois group of a polynomial $f(x)$?

Let us only focus on the case when $f(x)$ is irreducible. Then in the splitting field K of F, $f(x)$ factors as $f(x) = (x - \alpha_1) \cdots (x - \alpha_n)$. The Galois group $G = \mathrm{Gal}(K/F)$ acts on $\{\alpha_1, \dots, \alpha_n\}$. This then gives an embedding $G = \mathrm{Gal}(K/F) \hookrightarrow S_n$. The following is a simple observation.

Lemma 19.3.4. *The group G acts transitively on the set $\{\alpha_1, \dots, \alpha_n\}$.*

Proof. Suppose now and suppose that $\{\alpha_1, \dots, \alpha_r\}$ (with $r < n$) is an orbit under G, then $(x - \alpha_1) \cdots (x - \alpha_r) \in F[x]$ is a factor of $f(x)$. Yet $f(x)$ is irreducible; so this cannot happen. □

Before giving method of determining the Galois group, we study a "universal case".

Notation 19.3.5. Let F be a field, then we can consider the function field $M := F(x_1, \dots, x_n)$, namely

$$M = F(x_1, \dots, x_n) = \Big\{ \frac{p(\underline{x})}{q(\underline{x})} \,\Big|\, p(\underline{x}), q(\underline{x}) \in F[x_1, \dots, x_n], \ , q(\underline{x}) \neq 0 \Big\},$$

where $x_1, \dots, x_n$ are indeterminates. Define the **elementary symmetric functions** to be

$$s_1 = x_1 + \cdots + x_n, \quad s_2 = \sum_{i<j} x_i x_j, \quad \dots, \quad s_n = x_1 x_2 \cdots x_n.$$

Proposition 19.3.6. *The field $M = F(x_1, \dots, x_n)$ is a Galois extension over $L = F(s_1, \dots, s_n)$ with Galois group S_n.*

Proof. Consider the polynomial

$$f(x) = (x - x_1) \cdots (x - x_n) = x^n - s_1 x^{n-1} + \cdots + (-1)^n s_n \in L[x].$$

This M is the splitting field of $f(x)$ over L. In particular, as x_j's are distinct, M is separable and hence Galois over L.

Moreover, the Galois group $\mathrm{Gal}(M/L)$ permuting $\{x_1, \dots, x_n\}$ defines a homomorphism $\mathrm{Gal}(M/L) \hookrightarrow S_n$. But on the other hand, S_n acts on M as automorphism fixing L; so $S_n \subseteq \mathrm{Gal}(M/L)$. Thus, we have $\mathrm{Gal}(M/L) \cong S_n$. □

Slogan 19.3.7. The way we solve equations is modeled on the "universal function field" case.

The rest of this subsection aims to explain this slogan.

19.3.8. *Discriminant.* We first study the universal case $M = F(x_1, \dots, x_n)$ over $L = F(s_1, \dots, s_n)$. Consider the **discriminant**

$$\tilde{D} = \prod_{1 \le i < j \le n} (x_i - x_j)^2, \quad \text{and} \quad \sqrt{\tilde{D}} = \prod_{1 \le i < j \le n} (x_i - x_j).$$

We know that $\sigma \in S_n$ acts on M by sending $\sigma(\sqrt{\tilde{D}}) = \mathrm{sgn}(\sigma)\sqrt{\tilde{D}}$, where $\mathrm{sgn} : S_n \to \{\pm 1\}$ is the sign function with kernel $\ker(\mathrm{sgn}) = A_n$. (In fact, this is the definition of the sign function.)

This shows the following diagram of Galois extensions.

$$\begin{tikzcd}
M = F(x_1, \dots, x_n) \arrow[dd, dash, "S_n"'] \arrow[dr, dash, "A_n"] & \\
& L(\sqrt{\tilde{D}}) \arrow[dl, dash] \\
L = F(s_1, \dots, s_n) &
\end{tikzcd}$$

Now, we discuss the number field version. Let K be a splitting field of an irreducible polynomial $f(x) = (x - \alpha_1) \cdots (x - \alpha_n) \in F[x]$. Define the **discriminant** of f to be

$$D = \prod_{1 \le i < j \le n} (\alpha_i - \alpha_j)^2;$$

it belongs to F because it is invariant under all $\sigma \in \mathrm{Gal}(K/F)$.

Lemma 19.3.9. *We have $G = \mathrm{Gal}(K/F) \subseteq A_n$ if and only if D is a square in F. More precisely, we have the following diagram of Galois extensions:*

$$\begin{tikzcd}
K = F(\alpha_1, \ldots, \alpha_n) \arrow[dd, dash, "G"'] \arrow[dr, dash, "G\cap A_n"] & \\
& F(\sqrt{D}) \arrow[dl, dash] \\
F &
\end{tikzcd}$$

Proof. The second statement implies the first one because its shows that $F(\sqrt{D}) = F$ if and only if $G \subseteq A_n$.

Indeed, $\delta := \prod\limits_{1\leq i<j\leq n} (\alpha_i - \alpha_j) \in K$ is a square root of D, and $G \cap A_n$ clearly stabilizes δ and $G\backslash A_n$ does not. This implies that the Galois group of K over $F(\sqrt{D})$ is precisely $G \cap A_n$. □

19.3.10. *Solving cubic polynomials over $\mathbb{Q}$.* We apply the discussion above to (1) determine the Galois group of a cubic irreducible polynomial, and (2) give a general method to solve such a polynomial in $\mathbb{C}$ (by roots). By simple change of variables, we may assume that the polynomial is $x^3 + px + q \in \mathbb{Q}[x]$ (with zeros α, β, γ).

Similar to the above, we have the following diagram of intermediate fields.

$$\begin{tikzcd}
K = \mathbb{Q}(\alpha, \beta, \gamma) \arrow[dd, dash, "G"'] \arrow[dr, dash, "G\cap A_3"] & \\
& \mathbb{Q}(\sqrt{D}) \arrow[dl, dash] \\
\mathbb{Q} &
\end{tikzcd}$$

In this case, $D = (\alpha - \beta)^2(\beta - \gamma)^2(\gamma - \alpha)^2 = -4p^3 - 27q^2$. Combining this with the fact that $\mathrm{Gal}(K/\mathbb{Q})$ has to act transitively on the three roots, we deduce that

- if D is a square in $\mathbb{Q}$, then $\mathrm{Gal}(K/\mathbb{Q}) = A_3 = \mathbb{Z}_3$, and
- if D is not a square in $\mathbb{Q}$, then $\mathrm{Gal}(K/\mathbb{Q}) = S_3$.

To give a way to explicitly solve the equation $x^3 + px + q = 0$, we first of all compute $\sqrt{D}$. Next, we want to solve α, β, γ through understanding the cyclic extension. This goes back to the method of Kummer theory. We first adjoin $\omega := e^{2\pi\mathbf{i}/3}$ to the field so that we may use Kummer theory.

$$\begin{tikzcd}
& K(\omega) \arrow[dd, dash] \\
K = \mathbb{Q}(\alpha, \beta, \gamma) \arrow[ur, dash] \arrow[dd, dash] & \\
& \mathbb{Q}(\sqrt{D})(\omega) \\
\mathbb{Q}(\sqrt{D}) \arrow[ur, dash] &
\end{tikzcd}$$

Indeed, using the Lagrange resolvent presented in (19.1.3.1), we put

$$\theta_1 := \alpha + \omega\beta + \omega^2\gamma \quad \text{and} \quad \theta_2 := \alpha + \omega^2\beta + \omega\gamma.$$

Then we may directly compute

$$\theta_1^3 = \cdots = -\tfrac{27}{2}q + \tfrac{3}{2}\sqrt{-3D} \quad \text{and} \quad \theta_1^3 = \cdots = -\tfrac{27}{2}q - \tfrac{3}{2}\sqrt{-3D}.$$

From this, together with $\alpha + \beta + \gamma = 0$, we may solve α, β, and γ by linear algebra.

19.3.11. *Solving quartic polynomials over $\mathbb{Q}$.* We now turn to solve irreducible quartics and determine the associated Galois groups. Since S_4 is solvable, we consider the following chain of subgroups

$$1 \lhd V \lhd A_4 \lhd S_4,$$

where $V = \{1, (12)(34), (13)(24), (14)(23)\}$ is the Klein 4-group, and the subquotients of the above chain are all abelian groups. (We are slightly lucky here that V is a normal subgroup of S_4; so there is no ambiguity in how V is embedded in S_4.) Corresponding to this chain of extensions, we have

$$\begin{tikzcd}
M = F(x_1, \dots, x_4) \arrow[d, dash, "4"] & \{1\} \arrow[d, dash] \\
M^V = L(\sqrt{\tilde{D}}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3) \arrow[d, dash, "3"] & V \arrow[d, dash] \\
L(\sqrt{\tilde{D}}) \arrow[d, dash, "2"] & A_4 \arrow[d, dash] \\
L = F(s_1, \dots, s_4) & S_4,
\end{tikzcd}$$

where $\tilde{D}$ is the discriminant earlier and

$$\tilde{\theta}_1 = (x_1 + x_2)(x_3 + x_4), \quad \tilde{\theta}_2 = (x_1 + x_3)(x_2 + x_4), \quad \tilde{\theta}_3 = (x_1 + x_4)(x_2 + x_3).$$

Now, we consider the case of Galois extension of an irreducible quartic polynomial $x^4 + ax^2 + bx + c \in \mathbb{Q}[x]$, which has zeros $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ in the splitting field K. We look at a similar tower

$$\begin{tikzcd}
K \arrow[d, dash] \arrow[d, dash, bend left, "G\cap V"] \arrow[dd, dash, bend right=60, "G\cap A_4"'] \arrow[ddd, dash, bend left=70, "G"] \\
F(\sqrt{D})(\theta_1, \theta_2, \theta_3) \arrow[d, dash] \\
F(\sqrt{D}) \arrow[d, dash] \\
F
\end{tikzcd}$$

where $\sqrt{D} = \prod\limits_{1 \leq i < j \leq 4} (\alpha_i - \alpha_j)$ and

$$\theta_1 = (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4), \quad \theta_2 = (\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4), \quad \theta_3 = (\alpha_1 + \alpha_4)(\alpha_2 + \alpha_3).$$

(In principal), may use similar Kummer theory to solve this tower, as follows. First, a tedious computation shows that the discriminant is

$$D = \prod_{1 \leq i < j \leq 4} (\alpha_i - \alpha_j)^2 = 16a^4c - 3a^3b^2 - 128a^2c^2 + 144ab^2c - 12b^4 + 256c^3.$$

Then, to compute $\theta_1, \theta_2, \theta_3$, we adjoin the 3rd root of unity $\omega = e^{2\pi \mathbf{i}/3}$ and put

$$\eta_1 = \theta_1 + \omega\theta_2 + \omega^2\theta_3 \quad \text{and} \quad \eta_2 = \theta_1 + \omega^2\theta_2 + \theta_3.$$

It is expected that η_1^3 and η_2^3 can be expressed in terms of $\sqrt{D}$ by Kummer theory. This together with the fact that $\theta_1 + \theta_2 + \theta_3 = a$ allow us to solve for θ_1, θ_2, and θ_3. Here, we remark that, in addition to the notation symmetry, we note that the quotient group A_4/V is the Galois group in the universal case is generated by the permutation (132) (for example), it will take θ_1 to θ_2 and θ_2 to θ_3; so the above construction indeed follows the Kummer theory.

Next, we try to solve $\alpha_1, \dots, \alpha_4$ from knowing $\theta_1, \theta_2, \theta_3$. This corresponds to a biquadratic extension. For example, we may first understand $K^{(12)(34)}$; this is generated by $\alpha_1 + \alpha_2$ and $\alpha_3 + \alpha_4$. By we note that

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0 \quad \text{and} \quad (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4) = \theta_1.$$

This amounts to solve a quadratic equation. Similarly, we can get $\alpha_1\alpha_2$ and $\alpha_3\alpha_4$ by noting that

$$\alpha_1\alpha_2 + \alpha_3\alpha_4 = \tfrac{1}{2}(\theta_1 + \theta_3 - \theta_2) \quad \text{and} \quad \alpha_1\alpha_2\alpha_3\alpha_4 = c.$$

One can immediately solve α_1 and α_2 from these.

Next, we move to the study of the Galois group G. As the Galois group $G(K/F)$ can be viewed as a subfield of S_4 which acts transitively on all roots, it must belong to the following list:

$$S_4, \quad A_4, \quad V, \quad C = \langle (1234) \rangle, \quad \text{and} \quad D_8 = \langle (1234), (12)(34) \rangle$$

and their conjugates (only C and D_8 are not normal).

We list all possible intersections of these groups with V and with A_4.

$G =$	S_4	A_4	V	conjugates of C	conjugates of D_8
$\#G =$	24	12	4	4	8
$\#(A_4 \cap G) =$	12	12	4	2	8
$\#(A_4 \cap V) =$	4	4	4	2	4

So our computation will be able to determine the Galois group of the quartic equation.

20. Infinite Galois groups

20.1. A preliminary version of inverse limits.

Definition 20.1.1. Consider a sequence of surjective maps of sets

$$A_1 \overset{f_1}{\twoheadleftarrow} A_2 \overset{f_2}{\twoheadleftarrow} A_3 \overset{f_3}{\twoheadleftarrow} \cdots$$

Define

$$\varprojlim_n A_n := \big\{(a_1, a_2, \dots) \,\big|\, a_i \in A_i,\ f_i(a_{i+1}) = a_i\big\}.$$

This is called the **inverse limit**, **projective limit**, or just **limit** of the A_i's.

When each A_n has a structure of groups/rings, and f_n's are homomorphisms, the inverse limit is a group/ring.

Example 20.1.2. Let p be a prime number. Consider the inverse limit

$$\mathbb{Z}/p\mathbb{Z} \overset{f_1}{\twoheadleftarrow} \mathbb{Z}/p^2\mathbb{Z} \overset{f_2}{\twoheadleftarrow} \mathbb{Z}/p^3\mathbb{Z} \overset{f_3}{\twoheadleftarrow} \cdots$$

The inverse limit is $\mathbb{Z}_p := \varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$, the **ring of p-adic numbers** is

$$\mathbb{Z}_p := \big\{(x_1, x_2, \dots) \,\big|\, x_i \in \mathbb{Z}/p^i\mathbb{Z},\ x_{i+1} \bmod p^i = x_i\big\}.$$

The ring $\mathbb{Z}_p$ has very interesting properties. Take one example: for $p = 7$, we show that 2 is invertible in $\mathbb{Z}_7$ as follows:

$$\begin{aligned} 2x_1 &\equiv 1 \bmod 7 &&\Rightarrow && x_1 \equiv 1 \bmod 7 \\ 2x_2 &\equiv 1 \bmod 49 &&\Rightarrow && x_2 \equiv 25 \bmod 49\ (\equiv 4 \bmod 7) \\ &\cdots && && \cdots \end{aligned}$$

We can always solve $2x_i \equiv 1 \bmod 7^i$. This implies that 2 is invertible in $\mathbb{Z}_7$.

The same argument shows that

$$\mathbb{Z}_p^\times := \big\{(x_1, x_2, \dots) \in \mathbb{Z}_p \,\big|\, x_1 \neq 0\big\}.$$

We have $\mathbb{Z}_p = \mathbb{Z}_p^\times \sqcup p\mathbb{Z}_p$.

We have a natural map

$$\begin{aligned} \mathbb{Z} &\longrightarrow \mathbb{Z}_p = \varprojlim_n \mathbb{Z}/p^n\mathbb{Z} \\ a &\longmapsto \big(a \bmod p, a \bmod p^2, \dots\big). \end{aligned}$$

We can generalize the above construction as follows.

Lemma 20.1.3. *If (R_n, f_n) is an inverse system of rings and $R = \varprojlim_{n\to\infty} R_n$, then $R^\times = \varprojlim_{n\to\infty} R_n^\times$.*

Proof. By definition, we have

$$R^\times = \big\{\underline{a} = (a_1, a_2, \dots) \in R \,\big|\, \exists\, \underline{b} = (b_1, b_2, \dots) \in R, \text{ such that } \underline{a} \cdot \underline{b} = 1\big\}.$$

The condition implies that for each $\underline{a} \in R^\times$, $a_n \in R_n^\times$ for every n.

Conversely, for each $\underline{a} \in \varprojlim_{n\to\infty} R_n^\times$, the inverse a_n^{-1} of each a_n is unique, so certainly, we have

$$f_n(a_{n+1}^{-1}) = a_n^{-1}.$$

This implies that $(a_n^{-1})_n \in \varprojlim_{n\to\infty} R_n$ is the inverse of $\underline{a} \in R$. □

Example 20.1.4. Why did we call this a limit? We can see this by the following example:

$$\mathbb{C}[\![x]\!] := \varprojlim_n \mathbb{C}[x]/(x^n).$$

Here $\mathbb{C}[\![x]\!]$ is called the **ring of formal power series**. Given a smooth function f on near 0, the Taylor expansion at 0:

$$f(0) + xf'(0) + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots$$

defines an element in $\mathbb{C}[\![x]\!]$.

20.2. **A general inverse limit.**

Definition 20.2.1. A **partially ordered set (poset)** is a subset I such that for any two elements $i \neq j \in J$, either $i < j$, or $j < i$, or not comparable, satisfying

$$i < j,\ j < k\ \Rightarrow i < k.$$

We say that that I is **filtered** if for any $i, j \in I$, there exists $k \in I$ such that $i < k$ and $j < k$.

Definition 20.2.2. Let I be a partially ordered set, and let $(A_i)_{i\in I}$ be an **inverse system**, that is for each $i \in I$, we are given a set A_i; and if $j > i$, we have a map $\varphi_{ji} : A_j \to A_i$ such that if $k > j > i$, then $\varphi_{ki} = \varphi_{ji} \circ \varphi_{kj}$, i.e. the following diagram commutes

$$\begin{tikzcd} & A_j \arrow[dr, "\varphi_{ji}"] & \\ A_k \arrow[ur, "\varphi_{kj}"] \arrow[rr, "\varphi_{ki}"] & & A_i. \end{tikzcd}$$

Then we define the **inverse limit** to be

$$\varprojlim_{i\in I} A_i := \left\{(a_i)_{i\in I} \,\middle|\, a_i \in A_i,\ \text{and if } j > i,\ \text{we have } \varphi_{ji}(a_j) = a_i\right\} \subseteq \prod_{i\in I} A_i.$$

If each A_i is a group/ring and each φ_{ji} is a homomorphism, then $\varprojlim_{i\in I} A_i$ is a group/ring.

We have a natural map

$$\pi_i : \varprojlim_{i\in I} A_i \to A_i.$$

Fact 20.2.3. If B is a set/group/ring with maps/homomorphisms $\lambda_i : B \to A_i$ such that, for $j > i$, we have $\lambda_i = \varphi_{ji} \circ \lambda_j$, i.e. we hav the following commutative diagram

$$\begin{tikzcd} B \arrow[r, "\lambda_j"] \arrow[dr, "\lambda_i"'] & A_j \arrow[d, "\varphi_{ji}"] \\ & A_i, \end{tikzcd}$$

then we have a natural map $\lambda : B \to \varprojlim_{i\in I} A_i$.

Example 20.2.4. We define $\widehat{\mathbb{Z}} := \varprojlim_n \mathbb{Z}/n\mathbb{Z}$where the inverse system is given by divisibility (i.e. for $m|n$, $\mathbb{Z}/n\mathbb{Z} \twoheadrightarrow \mathbb{Z}/m\mathbb{Z}$).

It is fact that $\widehat{\mathbb{Z}} \cong \prod_{p \text{ prime}} \mathbb{Z}_p$.

We give the following proof: for each prime p, we have

$$\begin{aligned} \varphi_p : \widehat{\mathbb{Z}} &\longrightarrow \mathbb{Z}_p \\ (a_n)_n &\longmapsto (a_{p^r})_r. \end{aligned}$$

This together give a map

$$\varphi = \prod_p \varphi_p : \ \widehat{\mathbb{Z}} \to \prod_p \mathbb{Z}_p.$$

Conversely, to define a map $\prod_p \mathbb{Z}_p \to \widehat{\mathbb{Z}}$, it is enough to give a compatible family of, for each $n \in \mathbb{N}$,

$$\prod_p \mathbb{Z}_p \to \mathbb{Z}/n\mathbb{Z}.$$

For $n = p_1^{\alpha_1} \cdots p_s^{\alpha_s}$, this can be constructed as follows:

$$\prod_p \mathbb{Z}_p \to \mathbb{Z}_{p_1} \times \cdots \times \mathbb{Z}_{p^r} \twoheadrightarrow \mathbb{Z}/p_1^{\alpha_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_r^{\alpha_r}\mathbb{Z} \cong \mathbb{Z}/n\mathbb{Z}.$$

A similar example is that

$$\mathrm{GL}_N(\widehat{\mathbb{Z}}) \cong \prod_{p \text{ prime}} \mathrm{GL}_N(\mathbb{Z}_p).$$

20.3. Topology on inverse limit.

Definition 20.3.1. Let I be a *filtered* poset. Let $(A_i)_{i\in I}$ be an inverse system of sets. If each A_i carries a topology, we require the maps φ_{ji} to be continuous. (But for most of the case, we provide each A_i with discrete topology.)

We may define a topology on the inverse limit $A = \varprojlim_{i\in I} A_i$ as follows.

(1) (When each A_i is provided with discrete topology,) an open subset is the union of *basic opens*: for each $i \in I$ and each $a_i \in A_i$, the subset $\pi_i^{-1}(a_i) \subseteq A$ is a basic open subset, where $\pi_i : A \to A_i$ is the projection.

(2) (For general A_i,) embed

$$\varprojlim_{i\in I} A_i \subseteq \prod_{i\in I} A_i,$$

so that the right hand side is endowed with the product topology and the inverse limit is defined by the subspace topology.

Lemma 20.3.2. *When each A_i is provided discrete topology, the two above definitions of topology on $\varprojlim_{i\in I} A_i$ are equivalent.*

Proof. Clearly, an open subset in (1) is open in the sense of (2). Conversely, we note that an open subset as defined in (2) takes the form of $\pi_{i_1}^{-1}(a_{i_1}) \cap \cdots \cap \pi_{i_n}^{-1}(a_{i_n})$, where $i_1, \ldots, i_n \in I$ and $a_{i_j} \in A_{i_j}$. We need to show that such set is open in the sense of (1).

By induction, it is enough to show that for $i, j \in I$, $a_i \in A_i$, and $a_j \in A_j$, the intersection $\pi_i^{-1}(a_i) \cap \pi_j^{-1}(a_j)$ is open in topology defined in (1). As I is filtered, there exists $k \in I$ such that $i < k$ and $j < k$.

$$\begin{tikzcd}
A_i & \\
& A_k, \arrow[ul, "\varphi_{ki}"'] \arrow[dl, "\varphi_{kj}"] \\
A_j &
\end{tikzcd}$$

Put $B_k := \varphi_{ki}^{-1}(a_i) \cap \varphi_{kj}^{-1}(a_j)$. Then we have

$$\pi_i^{-1}(a_i) \cap \pi_j^{-1}(a_j) = \bigcup_{b \in B_k} \pi_k^{-1}(b).$$

The latter union is open in the topology defined in (1). □

Theorem 20.3.3. *If each A_i is finite (with discrete topology), then $\varprojlim_{i \in I} A_i$ is compact and Hausdorff. In this case, we say that $\varprojlim_{i \in I} A_i$* **profinite***.*

Proof. We consider the subspace

$$\varprojlim_{i \in I} A_i \subseteq \prod_{i \in I} A_i.$$

The latter space is compact and Hausdorff (as the product of compact spaces is compact by Tychonoff theorem). The subspace is determined by taking the conditions $\varphi_{ji}(a_j) = a_i$ for any $j > i$. This defined the left hand side as a closed subspace.

So $\varprojlim_{i \in I} A_i$ is compact and Hausdorff. □

Example 20.3.4. The topological rings $\mathbb{Z}_p$ and $\widehat{\mathbb{Z}}$ are profinite, and thus compact and Hausdorff. The topological groups $\mathbb{Z}_p^\times$ and $\widehat{\mathbb{Z}}^\times$ are also compact Hausdorff topological groups.

Definition 20.3.5. A **topological group** is a group G with a topology on the underlying subset such that the two maps

$$\begin{aligned} \iota : G &\longrightarrow G & \qquad m : G \times G &\longrightarrow G \\ g &\longmapsto g^{-1} & (g, h) &\longmapsto gh \end{aligned}$$

are continuous.

So if $U \subseteq G$ is an open subset, then $gUh \subseteq G$ is an open subset for any $g, h \in G$.

The following are interesting properties of topological groups.

Lemma 20.3.6. *If $H \leq G$ is an open subset of a topological group G, then H is also closed!*

Proof. Note that we have a disjoint union

$$G = \coprod_{gH \in G/H} gH$$

of subsets. But each gH is open. It implies that

$$H = G \backslash (\coprod_{gH \neq H} gH)$$

is closed. □

Lemma 20.3.7. *If G is a compact topological group, then a subgroup $H \leq G$ is open if and only if it is closed and of finite index in G.*

Proof. "$\Rightarrow$" Lemma 20.3.6 implies that H is closed. To see that H has finite index in G, we note that $G = \coprod gH$ is an open disjoint cover. Yet G is compact, so the number of subsets in the disjoint union is finite, i.e. $[G : H] < \infty$.

"$\Leftarrow$" If $H \leq G$ is a closed subgroup of finite index, we write $G = \coprod gH$, then we have

$$H = G \backslash (\bigcup_{gH \neq H} gH).$$

The union on the right hand side is a finite union, so the union is closed, and thus the complement H is open. □

Definition 20.3.8. A **profinite group** is an inverse limit of finite groups, with the inverse limit topology.

Lemma 20.3.9. *For a profinite group G, we have*

$$G := \varprojlim_{N \leq G \text{ open normal}} G/N.$$

(Note that all such N's form a filtered system: given $N_1 \lhd G$ and $N_2 \lhd G$, then $N_1 \cap N_2 \lhd G$.)

Proof. There is an obvious map $G \to \varprojlim_{N \lhd G \text{ open}} G/N =: G'$.

By definition, $G = \varprojlim_{i \in I} G_i$. We want to construct the reserve arrow:

$$G' = \varprojlim_{N \lhd G \text{ open}} G/N \to \varprojlim_{i \in I} G_i.$$

For this, it is enough to provide a compatible system of homomorphisms $G' \to G_i$ for each i. But note that the a natural map $\pi_i : G \to G_i$ has kernel $\ker \pi_i \lhd G$ (which has finite index since G_i is finite). Thus we can define the corresponding map

$$\lambda_i : G' = \varprojlim_{N \lhd G \text{ open}} G/N \twoheadrightarrow \frac{G}{\ker \pi_i} \to G_i.$$

This defines the a compatible system of maps $G' \to G_i$ (for each $i \in I$).

So far, we have not checked in any case the compatibility of the homomorphism (although the reader can keep in faith that they should be compatible.) Let us do this here: suppose that $i' > i$ is another index in I and we would like to check that the constructed maps $G' \to G_i$ and $G' \to G_{i'}$ are compatible in the sense that the following diagram commutes

$$\begin{tikzcd}
G' \arrow[r, "\lambda_{i'}"] \arrow[dr, "\lambda_i"'] & G_{i'} \arrow[d, "\varphi_{i'i}"] \\
& G_i
\end{tikzcd}$$

But this is clear because λ_i and $\lambda_{i'}$ expands to the following diagram

$$\begin{tikzcd}
G' \arrow[r, two heads] \arrow[dr, two heads] & \frac{G}{\ker \pi_{i'}} \arrow[r, "\cong"] \arrow[d] & G_{i'} \arrow[d, "\varphi_{i'i}"] \\
& \frac{G}{\ker \pi_i} \arrow[r, "\cong"] & G_i
\end{tikzcd}$$

The commutativity of the left triangle comes from the construct of the inverse limit G', and the commutativity of the right square comes from the inverse limit construction of G.

Now by Fact 20.2.3, we get a map $G' \to G$. By tracing back the definition, we see that this gives the inverse of the map $G \to G'$ we constructed above. □

Remark 20.3.10. In above lemma, we have written G as the inverse limit of G/N over *all* open normal subgroups N. In fact, we may also take the inverse limit over a subcollection of such N's, say $(N_j)_{j\in J}$ that is "cofinal" in the sense that for any open normal subgroup N of G, there exists some $N_j \le N$. Then we would have

$$G \cong \varprojlim_{j\in J} G/N_j.$$

This is because we certainly have a homomorphism $G \to \varprojlim_{j\in J} G/N_j$, to get the reverse arrow, it suffices to produce a compatible family of maps $\varprojlim_{j\in J} G/N_j \to G/N$. Indeed, by our cofinal condition, there exist some $N_{j_0} \le N$, we can therefore define

$$\lambda_N : \varprojlim_{j\in J} G/N_j \to G/N_{j_0} \to G/N. \tag{20.3.10.1}$$

Since the compatibility check here is a little more subtle (although the argument is still standard), we spell it out for the convenience of the readers. Rather the issue comes from the following: we need to first check that the definition of λ_N is independent of the choice of N_{j_0}. Indeed, if $N_{j_1} \le N$ is another open normal subgroup that is in the subcollection, we can then define $\lambda_{N,1} : G \to G/N$ similar to (20.3.10.1). We need to check that $\lambda_N = \lambda_{N,1}$. For this, we use the filtered property of J (in fact implied by the cofinal condition): there exists an open subgroup N_{j_2} in the collection such that $N_{j_2} \le N_{j_0} \cap N_{j_1}$. Now, we have the following commutative diagram

$$\begin{tikzcd}
& & G/N_{j_0} \arrow[dr] & \\
G = \varprojlim_{j\in J} G/N_j \arrow[r] \arrow[rrr, bend left=60, dashed, "\lambda_N"] \arrow[rrr, bend right=60, dashed, "\lambda_{N,1}"'] & G/N_{j_2} \arrow[ur] \arrow[dr] & & G/N \\
& & G/N_{j_1} \arrow[ur] &
\end{tikzcd}$$

The commutativity of the diamond on the right is the natural one, and thus, we deduce that $\lambda_N = \lambda_{N,1}$.

Now, we check that whenever $N' \leq N$, the map defined above is compatible, namely, we need to check the commutativity of the following diagram.

$$\begin{tikzcd}
\varprojlim_{j\in J} G/N_j \arrow[r, two heads] \arrow[dr, two heads] & G/N_{j'_0} \arrow[r] & G/N' \arrow[d, "\varphi_{N'N}"] \\
 & G/N_{j_0} \arrow[r] & G/N
\end{tikzcd}$$

Here, $N_{j'_0}$ is an open subgroup in the subcollection such that $N_{j'_0} \leq N'$. A small subtlety here is that we do not know how to compare $N_{j'_0}$ with N_{j_0} in general. But we have just discussed earlier that we may replace the definition of $\lambda_{N'}$ using a smaller open subgroup $N_{j'_1}$ that is contained in N_{j_0}. This way, we get an arrow downwards $G/N_{j'_1} \to G/N_{j_0}$ and proves that $\lambda_N = \varphi_{N'N} \circ \lambda_{N'}$.

20.4. **Infinite Galois theory.** Let us recall that a Galois extension K over F is an extension that is separable and normal, i.e.

(1) any intermediate field E that is finite over F is separable over F,
(2) any irreducible polynomial $f(x) \in F[x]$ having one root in K splits over $K[x]$.

Remark 20.4.1. Such a field K is the union of intermediate field E such that E/F is finite and Galois.

Definition 20.4.2. Let K be a Galois extension of F. Define its Galois group to be

$$\mathrm{Gal}(K/F) := \varprojlim_{E/F \text{ finite Galois}} \mathrm{Gal}(E/F).$$

The connecting map is, if $E_1 \supseteq E_2 \supset F$ are finite extensions, we have $\mathrm{Gal}(E_1/F) \twoheadrightarrow \mathrm{Gal}(E_2/F)$. Note that this defines $\mathrm{Gal}(K/F)$ as a profinite group (with topology).

Example 20.4.3. (1) Write $\mathbb{Q}(\mu_{p^\infty}) := \mathbb{Q}(\zeta_{p^n}; n \in \mathbb{N})$. We have

$$\mathrm{Gal}\big(\mathbb{Q}(\mu_{p^\infty})/\mathbb{Q}\big) = \varprojlim_n \mathrm{Gal}\big(\mathbb{Q}(\zeta_{p^n})/\mathbb{Q}\big) \cong \varprojlim_n \big(\mathbb{Z}/p^n\mathbb{Z}\big)^\times = \mathbb{Z}_p^\times.$$

(2) Write $\mathbb{Q}(\mu_\infty) := \mathbb{Q}(\zeta_n; n \in \mathbb{N})$. We have

$$\mathrm{Gal}\big(\mathbb{Q}(\mu_\infty)/\mathbb{Q}\big) = \varprojlim_n \mathrm{Gal}\big(\mathbb{Q}(\zeta_n)/\mathbb{Q}\big) \cong \varprojlim_n \big(\mathbb{Z}/n\mathbb{Z}\big)^\times = \hat{\mathbb{Z}}^\times \cong \prod_p \mathbb{Z}_p^\times.$$

Lemma 20.4.4. *If K is a Galois extension of F, then the Galois group $\mathrm{Gal}(K/F)$ as an abstract group is isomorphic to*

$$\mathrm{Gal}(K/F) \cong \big\{\textit{automorphism } \sigma : K \xrightarrow{\simeq} K \;\big|\; \sigma|_F = \mathrm{id}_F\big\}. \tag{20.4.4.1}$$

Proof. Giving an automorphism $\sigma : K \xrightarrow{\simeq} K$ such that $\sigma|_F = \mathrm{id}_F$ is equivalent to giving, for any finite normal intermediate field E/F, a compatible automorphism $\sigma_E : E \xrightarrow{\simeq} E$ such that $\sigma_E|_F = \mathrm{id}_F$. This is precisely the definition of $\mathrm{Gal}(K/F)$ in Definition 20.4.2. □

Remark 20.4.5. By Lemma 20.4.4, Gal(K/F) is a profinite group; its topology is given by the following: if E is an intermediate finite Galois extension of F and if $a_E \in \mathrm{Gal}(E/F)$ is an element, then

$$\big\{\sigma : K \xrightarrow{\simeq} K \,\big|\, \sigma|_E = a_E\big\}$$

is a *standard* open subsets of Gal(K/F).

The open subsets of Gal(K/F) are unions of such standard opens.

Remark 20.4.6. The Galois group Gal(K/F) acts on K continuously, this is to say that the "action map"

$$\mathrm{act} : \mathrm{Gal}(K/F) \times K \to K$$

is continuous when K is endowed with discrete topology.

Indeed, to verify the continuity of act, we take any open subset, or rather just any element $\alpha \in K$, and consider the preimage $\mathrm{act}^{-1}(\alpha)$. This preimage consists of tuples $(g, x) \in \mathrm{Gal}(K/F) \times K$ such that $g(x) = \alpha$. Note that the conjugates of α in K is finite, say $\alpha_i = \sigma_i(\alpha)$ for $i = 1, \dots, r$ with $\sigma_1 = \mathrm{id}$. Then explicitly,

$$\mathrm{act}^{-1}(\alpha) = \bigsqcup_{i=1}^{r} \Big(\mathrm{Gal}(K/F(\alpha))\sigma_i^{-1} \times \{\alpha_i\}\Big);$$

it is a finite union of open subsets.

Theorem 20.4.7 (Galois theory for infinite extensions)**.** *Let K be a Galois extension of F. Then there is a one-to-one inclusive-reversing correspondence*

$$\big\{\textit{closed subgroups } H \textit{ of } \mathrm{Gal}(K/F)\big\} \;\longleftrightarrow\; \big\{\textit{intermediate fields } L \textit{ of } K/F\big\}.$$

Moreover, we have the following.

(1) *If $H \longleftrightarrow L$, then H is open if and only if L/F is a finite extension.*

(2) *If $H_1, H_2 \;\longleftrightarrow\; L_1, L_2$, then*

$$H_1 \cap H_2 \;\longleftrightarrow\; L_1L_2 \quad \textit{and} \quad \langle H_1, H_2\rangle \;\longleftrightarrow\; L_1 \cap L_2.$$

(3) *If closed subgroups $H_1 \le H_2$ corresponds to $L_1 \ge L_2$, then*

$$[H_2 : H_1] = [L_1 : L_2].$$

(4) *If $H \longleftrightarrow L$ and $g \in \mathrm{Gal}(K/F)$, then*

$$gHg^{-1} \longleftrightarrow g(L).$$

(5) *If $H \longleftrightarrow L$, then H is a normal subgroup if and only if L is a normal extension of F. In this case,*

$$\mathrm{Gal}(L/F) \cong G/H$$

as topological groups*, where we equip G/H with the quotient topology, i.e. for the natural projection $\pi : G \to G/H$, a subset $V \subseteq G/H$ is open if and only if $\pi^{-1}(V)$ is open in G. (Note that G/H is also a profinite group.)*

Proof. We will only prove the correspondence part, especially explaining the closedness condition on the subgroup H. The rest part of the theorem should mostly follow from the same argument as before, except (3), which we will make some remarks at the end. Before giving the proof, we investigate the profinite inverse limit further.

(A) (Profinite group side) Let G be a profinite group; so $G = \varprojlim_{N \lhd G \text{ open}} G/N$ by Lemma 20.3.9. Let H be a closed subgroup. For each $N \lhd G$ open normal, define

$$H_N := \mathrm{Im}(H \to G/N).$$

Then for another open normal subgroup $N' \lhd G$ such that $N' \leq N$, we have a natural map $H_{N'} \to H_N$. We use the following big diagram to represent the relevant groups.

$$\begin{tikzcd}
G \arrow[d] & H \arrow[l] \arrow[d] \\
\vdots \arrow[d] & \vdots \arrow[d] \\
G/N' \arrow[d] & H_{N'} := \mathrm{Im}(H \to G/N') \arrow[l] \arrow[d] \\
G/N & H_N := \mathrm{Im}(H \to G/N). \arrow[l]
\end{tikzcd}$$

We claim that H, as a subgroup, is isomorphic to

$$H \cong \varprojlim_{N \lhd G \text{ open}} H_N \subseteq G. \tag{20.4.7.1}$$

(We note that $\varprojlim_{N \lhd G \text{ open}} H_N = G \cap \Big(\bigcap_{N \lhd G \text{ open}} \pi_N^{-1}(H_N) \Big)$ is a closed subgroup (where $\pi_N : G \to G/N$ is the projection); so we really need H to be a closed subgroup to start.)

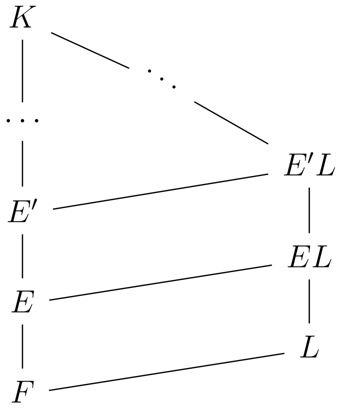
Proof of the claim: Suppose that the inclusion $H \subseteq \varprojlim_{N \lhd G \text{ open}} H_N$ is strict, then writing H^c for the complement,

$$H^c \cap \Big(\varprojlim_{N \lhd G \text{ open}} H_N \Big) \neq \emptyset.$$

So there exists a basic open subgroup $gN_0 \subseteq H^c$ (for some open $N_0 \lhd G$) such that $gN_0 \cap H = \emptyset$ but $gN_0 \cap \Big(\varprojlim_{N \lhd G \text{ open}} H_N \Big) \neq \emptyset$.

The first condition implies that $g \notin H_N$, yet the second condition implies that $g \in H_N$. This is a contradiction! So the isomorphism (20.4.7.1) holds.

(B) (Field extension side) Let L be an intermediate field of K/F. Consider the following diagram of fields (where E and E' are intermediate fields of K/F that are finite and Galois over F.



Recall that we have an inverse limit

$$\mathrm{Gal}(K/F) := \varprojlim_{E/F \text{ finite Galois}} \mathrm{Gal}(E/F). \tag{20.4.7.2}$$

For an intermediate field L of K/F,

$$\mathrm{Gal}(K/L) := \varprojlim_{L'/L \text{ finite Galois}} \mathrm{Gal}(L'/L).$$

Note that for each such L', we may find a finite Galois extension E/F such that $L' \subseteq EL$ (and automatically EL is Galois over L). So finite Galois extensions of L of the form EL form a final system. This implies that

$$\mathrm{Gal}(K/L) \cong \varprojlim_{E/F \text{ finite Galois}} \mathrm{Gal}(EL/L) \overset{\text{Prop 18.2.1}}{\cong} \varprojlim_{E/F \text{ finite Galois}} \mathrm{Gal}(E/(L \cap E)). \tag{20.4.7.3}$$

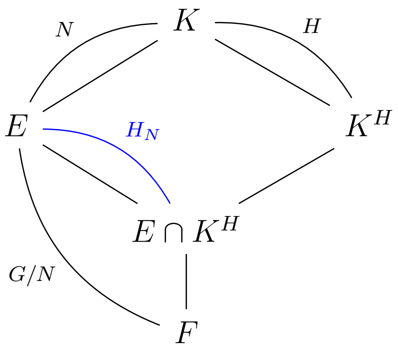
As each $\mathrm{Gal}(E/(L \cap E))$ is a subgroup of $\mathrm{Gal}(E/F)$, this in particularly implies that $\mathrm{Gal}(K/L)$ is naturally a closed subgroup of $\mathrm{Gal}(K/F)$.

Now, we are in position to prove the main correspondence.

First, we check that, for a closed subgroup $H \leq G$, $\mathrm{Gal}(K/K^H) = H$. As discussed in (B), we have

$$\mathrm{Gal}(K/K^H) = \varprojlim_{E/F \text{ finite Galois}} \mathrm{Gal}(E/(E \cap K^H)).$$

For each E/F finite Galois, $N = \mathrm{Gal}(K/E)$ is an open normal subgroup of G; so $\mathrm{Gal}(E/F) \cong G/N$. We have the following diagram of fields.



Here, H acts on the field E via the restriction map

$$\begin{tikzcd}
\mathrm{Gal}(K/F) \arrow[r] \arrow[d, equal] & \mathrm{Gal}(E/F) \arrow[d, equal] \\
H \arrow[r] & G/N
\end{tikzcd}$$

So $E \cap K^H = E^{H_N}$. This implies that

$$\mathrm{Gal}(K/K^H) = \varprojlim_{E/F \text{ finite Galois}} \mathrm{Gal}\big(E/(E \cap K^H)\big) = \varprojlim_{E/F \text{ finite Galois}} \mathrm{Gal}(E/E^{H_N}) \cong \varprojlim_{E/F \text{ finite Galois}} H_N \overset{(A)}{=} H.$$

Here the second last isomorphism uses the finite Galois theory.

Secondly, we need to check that for any intermediate field L of K/F, $L = K^{\mathrm{Gal}(K/L)}$. For any finite Galois extension E/F, we need to check that

$$E \cap L = K^{\mathrm{Gal}(K/L)} \cap E.$$

But we know that

$$E \cap K^{\mathrm{Gal}(K/L)} = E^{\mathrm{Im}(\mathrm{Gal}(K/L) \to \mathrm{Gal}(E/F)} \overset{(B)}{=} E^{\mathrm{Gal}(E/E \cap L)} = E \cap L.$$

Finally, we give a few remarks on (3). We will only discuss the case when $[H_2 : H_1]$ and $[L_1 : L_2]$ are both finite. A modification of the argument handles the infinite case. There are two aspects here that correspond to each other under the Galois correspondence.

(Group aspect) For two closed subgroups $H_1 \leq H_2$, and an open normal subgroup N of a profinite group G, we may consider $NH_1 \leq NH_2$; their images in G/N defines subgroups $H_{1,N} \leq H_{2,N}$. When $N \leq N'$, there is a natural map of cosets:

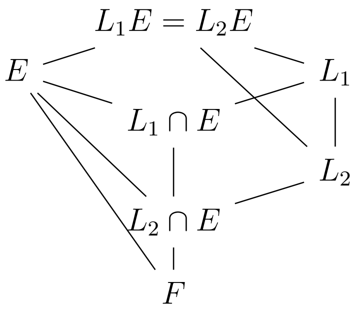
$$H_{2,N}/H_{1,N} \cong H_2N/H_1N \to H_2N'/H_1N' \cong H_{2,N'}/H_{1,N'}.$$

This map is clearly surjective. The inverse limit of $H_{2,N}/H_{1,N}$ along such connecting map is precisely H_2/H_1. In the special case when $[H_2 : H_1]$ is finite, for N sufficiently small, the natural map of cosets

$$H_2/H_1 \to NH_2/NH_1 \tag{20.4.7.4}$$

is a bijection.

(Field aspect) We have two intermediate extensions $K/L_1/L_2/F$. In general, for another intermediate extension E finite Galois over F, the field extension $L_1 \cap E$ over $L_2 \cap E$ can be a "very small" extension. But we will take E very "large". More precisely, when L_1/L_2 is a finite (separable) extension, then it is generated by one element α with minimal polynomial $m_{\alpha,L_2}(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_0$. Then for any E containing the Galois closure of $F(a_0, \ldots, a_{d-1}, \alpha)$ over F inside K, we must have $L_1E = L_2E$. Consider the following diagram



As E/F is finite and Galois, Proposition 18.2.1 implies that

$$[L_1E : L_1] = [E : L_1 \cap E] \quad \text{and} \quad [L_2E : L_2] = [E : L_2 \cap E].$$

It then follows that

$$[L_1 : L_2] = \frac{[L_2E : L_2]}{[L_1E : L_1]} = \frac{[E : L_2 \cap E]}{[E : L_1 \cap E]} = [L_1 \cap E : L_2 \cap E].$$

So based on the above discussion, if we take an open normal subgroup $N \lhd G$ such that (20.4.7.4) holds and that N fixes the field E in the (Field aspect) so that $L_1E = L_2E$. Then we have

$$[H_2 : H_1] = [NH_2 : NH_1] = [H_{2,N} : H_{1,N}] \quad \text{and} \quad [L_1 : L_2] = [L_1 \cap E : L_2 \cap E].$$

Now consider the finite Galois extension E/F, we have

$$[L_1 \cap E : L_2 \cap E] = [\mathrm{Gal}(E/E \cap L_2) : \mathrm{Gal}(E/E \cap L_1)] = [H_{2,N} : H_{1,N}].$$

This completes the proof of (3) (at least when the two quantities are finite). □

20.5. **Galois representation.**

Definition 20.5.1. For a group G, an n-dimensional **representation** over a field L is a homomorphism

$$\rho : G \to \mathrm{GL}_n(L).$$

This is equivalent to having G acting L-linearly on a n-dimensional vector space V over L; namely,

$$\rho(g)(av) = a\rho(g)(v) \qquad \text{for } a \in L.$$

It is a general philosophy that, to understand a group, it is "equivalent" to understand all of its representations. For profinite groups, we have the following.

Proposition 20.5.2. *If G is a profinite group, then any continuous representation $\rho : G \to \mathrm{GL}_n(\mathbb{C})$ has finite image.*

Proof. Take a very small open neighborhood U of $I_n \in \mathrm{GL}_n(\mathbb{C})$. The preimage $\rho^{-1}(U)$ is an open subset of G containing I_n. This implies that $\rho^{-1}(U)$ contains an open subgroup H of G.

Note now that $\rho(H) \subset U$. But if U is a small neighborhood of $I_n \in \mathrm{GL}_n(\mathbb{C})$, it cannot contain a subgroup of $\mathrm{GL}_n(\mathbb{C})$. This implies that $\rho(H) = \{I_n\}$. Thus ρ factors through G/H. □

So in order to study more interesting representations, one needs to consider representations of the kind $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{Q}_p)$, where $\overline{\mathbb{Q}}$ is the algebraic closure of $\mathbb{Q}$ inside $\mathbb{C}$. This general setting is the main object of the so-called Langlands program.

21. Algebraic closures and transcendent extensions

21.1. Algebraic closure and separable closure.

Definition 21.1.1. A field extension K of F is called an **algebraic closure** if

(1) K/F is an algebraic extension;
(2) every polynomial $f(x) \in F[x]$ splits completely over K.

A field extension K of F is called a **separable closure** if

(1) K/F is an algebraic separable extension;
(2) every separable polynomial $f(x) \in F[x]$ splits completely over K.

Typically, we write F^{alg} or $\overline{F}$ for an algebraic closure, and F^{sep} for a separable closure.

Note that we have not discussed the existence of algebraic or separable closures, nor the uniqueness of algebraic or separable closures yet. We will soon get to these topics.

Remark 21.1.2. If E/F is the splitting field of some polynomial $f(x) \in F[x]$, then by the claim in Proposition 15.2.4, there exists an embedding $E \hookrightarrow F^{\mathrm{alg}}$.

If E/F is the splitting field of a separable polynomial $f(x) \in F[x]$, then there exists an embedding $E \hookrightarrow F^{\mathrm{sep}}$.

So an algebraic closure contains any splitting field of F.

Definition 21.1.3. A field K is called **algebraically closed** if all polynomials in $K[x]$ splits completely. This is equivalent to that the only irreducible polynomial in $K[x]$ are linear (or constant) ones; which is in turn equivalent to that K has no nontrivial algebraic extension.

A field K is called **separably closed** if all nontrivial algebraic extensions are inseparable.

Proposition 21.1.4. (1) *An algebraic closure of an algebraically closed field K is just K.*
(2) *A separable closure of a separably closed field K is just K.*
(3) *If $\overline{F}$ is an algebraic closure of F, then $\overline{F}$ is algebraically closed.*

Proof. (1) This is because if $\alpha \in K^{\mathrm{alg}}$ is the zero of a polynomial in $K[x]$, then $\alpha \in K$ because K is algebraically closed. Thus, $K^{\mathrm{alg}} = K$.

(2) This is the same as (1).

(3) Suppose that α is algebraic over $\overline{F}$. We want to show that $\alpha \in \overline{F}$. Consider the minimal polynomial of α over $\overline{F}$:

$$m_{\alpha,\overline{F}}(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \overline{F}[x].$$

Consider the tower of extension

$$\begin{tikzcd}
F(\alpha, a_0, \ldots, a_{n-1}) \arrow[d, dash] \arrow[d, bend right, dash, "\text{finite}"'] \\
F(a_0, \ldots, a_{n-1}) \arrow[d, dash] \arrow[d, bend right, dash, "\text{finite}"'] \\
F
\end{tikzcd}$$

This implies that $F(\alpha, a_0, \ldots, a_{n-1})/F$ is a finite extension. So α is the zero of a polynomial in F; so $\alpha \in \overline{F}$. □

Theorem 21.1.5. (1) *Any field F is contained in an algebraically closed field K.*

(2) *If K/F is a field with K algebraically closed, then*
- $F^{\mathrm{alg}} := \{x \in K \mid x$ *algebraic over* $F\}$ *is an algebraic closure of F;*
- $F^{\mathrm{sep}} := \{x \in K \mid x$ *algebraic and separable over* $F\}$ *is a separable closure of F.*

(3) *The algebraic closure and separable closure of a field F are unique up to isomorphisms (but not up to canonical isomorphisms).*

Proof. (1) See the extended reading material after this lecture.

(2) By definition, F^{alg} is algebraic over F. Each polynomial $f(x) \in F[x]$ splits over K; yet the zeros are algebraic over F; so the zeros of $f(x)$ belongs to F^{alg}. Thus, $f(x)$ splits over F^{alg}.

Moreover, the same argument shows that every separable polynomial $f(x) \in F^{\mathrm{sep}}[x]$.

(3) If $F \hookrightarrow F^{\mathrm{alg}}$ and $F \hookrightarrow F^{\mathrm{alg}\prime}$ are two embeddings, then Proposition 15.2.4 implies that we have a natural (injective) homomorphism $\eta : F^{\mathrm{alg}} \to F^{\mathrm{alg}\prime}$. Conversely, the subfield $\eta(F^{\mathrm{alg}}) \subseteq F^{\mathrm{alg}\prime}$ gives an algebraic extension; so $\eta(F^{\mathrm{alg}}) = F^{\mathrm{alg}\prime}$ (as there is no nontrivial algebraic extension of $\eta(F^{\mathrm{alg}})$). This shows that F^{alg} is isomorphic to $F^{\mathrm{alg}\prime}$.

The same argument works for separable closures. □

21.2. **Transcendent extensions.** So far, we have been focusing on algebraic extensions. Now we come to studying "larger" extensions. Recall that in a field extension K/F, an element $\alpha \in K$ is called *transcendental* over F if $F[x] \to K$ sending x to α is an injection.

Definition 21.2.1. (1) Let K/F be a field extension. A subset $\{\alpha_1, \dots, \alpha_n\} \subseteq K$ is called **algebraically independent** over F, if there is no nonzero polynomial $f(x_1, \dots, x_n) \in F[x_1, \dots, x_n]$ such that $f(\alpha_1, \dots, \alpha_n) = 0$.

This gives rise to an injective homomorphism

$$\begin{aligned} \eta : F(x_1, \dots, x_n) &\longrightarrow K \\ p(\underline{x})/q(\underline{x}) &\longmapsto p(\underline{\alpha})/q(\underline{\alpha}). \end{aligned}$$

An infinite subset A of K is called **algebraically independent** over F if any finite subset of A is algebraically independent over F.

(2) A **transcendence generator subset** for a field extension K/F is a subset $A \subset K$ such that K is algebraic over $F(A)$. (Note that this name is not standard; or rather, this notation is rarely used in the literature.)

(3) A **transcendence base** for K/F is a subset $A \subset K$ that is algebraically independent and also a transcendence generator. This is equivalent to say A is a "maximal subset of K that is algebraically independent over F.

Remark 21.2.2. For many proofs below, one may make the following analogy between field extensions K/F with vectors spaces over F:

Field extensions	$\longleftrightarrow$	Vector spaces
Algebraically independent subsets	$\longleftrightarrow$	Linearly independent subsets
Transcendence generator subsets	$\longleftrightarrow$	Generating subsets
Transcendence bases	$\longleftrightarrow$	bases

Theorem 21.2.3. *Any field extension K/F has a transcendence base and any two transcendence bases of K/F have the same cardinality.*

Proof. The existence of transcendence bases follows from Zorn's lemma.

We need to show that the cardinality of an algebraic independent set is always smaller than or equal to that of a transcendence generator subset. We first treat the finite case, with an algebraically independent subset $\{\alpha_1, \dots, \alpha_m\}$ and a transcendence generator subset $\{\beta_1, \dots, \beta_n\}$.

If $\{\alpha_1, \dots, \alpha_m\} \subseteq \{\beta_1, \dots, \beta_n\}$, we have $m \leq n$, and we are done. Suppose not. WLOG, we assume that $\alpha_1 \notin \{\beta_1, \dots, \beta_n\}$. Since $\{\beta_1, \dots, \beta_n\}$ is a transcendence generator subset, $F(\alpha_1, \beta_1, \dots, \beta_n)$ must be algebraic over $F(\beta_1, \dots, \beta_n)$. Consider the minimal polynomial $m_{\alpha_1, F(\beta_1, \dots, \beta_n)}(x)$. Clearing its denominators, we obtain an (irreducible) polynomial $f(x, y_1, \dots, y_n) \in F[x, y_1, \dots, y_n]$ such that $f(\alpha_1, \beta_1, \dots, \beta_n) = 0$. We claim that there is some j (WLOG $j = 1$) such that

- $\beta_j \notin \{\alpha_1, \dots, \alpha_m\}$, and
- y_j appears in some term in $f(x, y_1, \dots, y_n)$.

Otherwise, the equality $f(\alpha_1, \beta_1, \dots, \beta_n) = 0$ is entirely algebraic relations among elements in $\{\alpha_1, \dots, \alpha_m\}$, contradicting that α_i's form an algebraically independent subset.

Now, consider the following diagram

$$\begin{tikzcd}
 & K \arrow[d, dash] & \\
 & F(\alpha_1, \beta_1, \dots, \beta_n) \arrow[dl, dash] \arrow[dr, dash, "\text{finite}"] & \\
F(\beta_1, \dots, \beta_n) \arrow[uur, dash, bend left, "\text{algebraic}"] & & F(\alpha_1, \beta_2, \dots, \beta_n).
\end{tikzcd}$$

The field extension $F(\alpha_1, \beta_1, \dots, \beta_n)/F(\alpha_1, \beta_2, \dots, \beta_n)$ is finite because β_1 satisfies a nontrivial relation $f(\alpha_1, \beta_1, \dots, \beta_n) = 0$. It then follows from this that K is algebraic over $F(\alpha_1, \beta_2, \dots, \beta_n)$, i.e. $\{\alpha_1, \beta_2, \dots, \beta_n\}$ is a transcendence generator subset.

Continuing this way, each time we swap one element α_1 into the set $\{\beta_1, \dots, \beta_n\}$, and eventually, we get $\{\alpha_1, \dots, \alpha_m\} \subseteq \{\beta_1, \dots, \beta_n\}$. Thus, $m \leq n$.

The infinite case is more subtle, using again the Zorn's lemma. We omit the details. □

Definition 21.2.4. The cardinality of a transcendence base for a field extension K/F is called the **transcendence degree** for K/F.

Remark 21.2.5. The field $\mathbb{Q}(\pi)$ is isomorphic to $\mathbb{Q}(t)$, which has transcendence degree 1.

Caveat 21.2.6. If $\{\alpha_1, \dots, \alpha_n\}$ and $\{\alpha'_1, \dots, \alpha'_n\}$ are transcendence bases, in general, the fields $F(\alpha_1, \dots, \alpha_n)$ and $F(\alpha'_1, \dots, \alpha'_n)$ may not be the same.

For example, x and x^2 are both transcendence bases for the extension $\mathbb{Q}(x)/\mathbb{Q}$. But $\mathbb{Q}(x^2) \neq \mathbb{Q}(x)$.

Proposition 21.2.7. *Let t be a transcendental variable over F. If $p(t), q(t) \in F[t]$ are relatively prime polynomials that are not both constant, then*

$$\Big[F(t) : F\Big(\frac{p(t)}{q(t)}\Big)\Big] = \max\big(\deg p(t), \deg q(t)\big).$$

Proof. Writing $y = \dfrac{p(t)}{q(t)}$, the minimal polynomial of t over $F(y)$ is

$$p(t) - yq(t) = 0, \quad \text{or} \quad q(t) - \frac{1}{y}p(t) = 0,$$

(depending on which of $p(t)$ and $q(t)$ has larger degree). The Proposition follows. □

Definition 21.2.8. An extension K/F is called **purely transcendent** if $K = F(\alpha_1, \dots, \alpha_n)$.

Question 21.2.9. How to describe a general K/F? What about its integral version? Can we "visualize" this?

For example, we have the following extension.

$$\begin{tikzcd}
F(x)(\sqrt{x^3+x}) \arrow[d, dash] & \supseteq & F[x,y]/(y^2 - x^3 - x) \arrow[d, dash] \\
F(x) & \supseteq & F[x].
\end{tikzcd}$$

The general case might be trickier.

21.3. **Basic ideas of algebraic geometry.** A basic idea of algebraic geometry is to study the relation between the space and the functions on such spaces, in the following sense:

$$\begin{aligned}
U \subseteq \mathbb{C}^n \text{ open} &\longleftrightarrow \mathcal{O}(U) := \{\text{holomorphic functions on } U\} \\
x \in U &\rightsquigarrow \mathfrak{m}_x := \{f \in \mathcal{O}(U) \mid f(x) = 0\} \\
&\qquad \mathcal{O}(U)/\mathfrak{m}_x \cong \mathbb{C}, \text{ so } \mathfrak{m}_x \text{ is maximal}
\end{aligned}$$

In the following, let k be an algebraically closed field (e.g. $k = \mathbb{C}$). The general philosophy is that there is a correspondence

$$\text{space } k^n \quad \longleftrightarrow \quad \text{polynomial ring } k[x_1, \dots, x_n].$$

This makes sense because a polynomial can be evaluated at every point $(a_1, \dots, a_n) \in k^n$. Moreover, each point $\underline{a} = (a_1, \dots, a_n) \in k^n$ corresponds to the maximal ideal $\mathfrak{m}_{\underline{a}} = (x_1 - a_1, \dots, x_n - a_n)$.

One of the most important theorems in algebraic geometry is the following.

Theorem 21.3.1 (Hilbert's Nullstellensatz, weak form)**.** *Assume that k is algebraically closed. Then every maximal ideal of $k[x_1, \dots, x_n]$ is of the form $(x_1 - a_1, \dots, x_n - a_n)$ for some $(a_1, \dots, a_n) \in k^n$. In other words, there is a one-to-one correspondence*

$$k^n \quad \longleftrightarrow \quad \big\{\textit{maximal ideals of } k[x_1, \dots, x_n]\big\}.$$

We will also discuss what happens if k is not algebraically closed in the next lecture.

To proceed, we need to discuss some commutative algebra.

Definition 21.3.2. Let R be a commutative ring and I and ideal. Define the **radical** of I to be the ideal

$$\sqrt{I} := \{f \in R \mid f^n \in I \text{ for some } n \in \mathbb{N}\}.$$

Obviously, $I \subseteq \sqrt{I}$ and $\sqrt{\sqrt{I}} = \sqrt{I}$. We say that I is **radical** if $I = \sqrt{I}$.

We need to check that $\sqrt{I}$ is indeed an ideal:

- for $f, g \in \sqrt{I}$, $f^m, g^n \in I$ for some $m, n \in \mathbb{N}$, then $(f+g)^{m+n-1} = f^m \cdot a + g^n \cdot b$ for some $a, b \in R$ (which are some polynomial expressions in f and g), thus $f + g \in \sqrt{I}$;
- for any $a \in R$, $f \in \sqrt{I}$, we have $f^n \in I$ for some $n \in \mathbb{N}$, and thus $(af)^n = a^n f^n \in I$; so $af \in \sqrt{I}$.

Lemma 21.3.3. *If $\mathfrak{p}$ is a prime ideal such that $I \subseteq \mathfrak{p}$, then $\sqrt{I} \subseteq \mathfrak{p}$. (In particular, this holds for maximal ideals as well.)*

Proof. If $a \in \sqrt{I}$, then $a^n \in I$ for some $n \in \mathbb{N}$; so $a^n \in \mathfrak{p}$, which implies that $a \in \mathfrak{p}$ as $\mathfrak{p}$ is a prime ideal. Thus, $\sqrt{I} \subseteq \mathfrak{p}$. □

Here is the picture for algebraic geometry: studying relations between "appropriate subsets of k^n" and rings that come from algebraic structures of $k[x_1, \dots, x_n]$.

$$\begin{tikzcd}
\text{subsets of } k^n \arrow[r, leftrightarrow] & \text{ideals of } k[x_1, \dots, x_n] \\
Z \arrow[r, mapsto] & I(Z) := \{f \in k[\underline{x}] \mid f(z) = 0, \forall z \in Z\} \\
Z(f) := \{\underline{a} \in k^n \mid f(\underline{a}) = 0\} & (f) \arrow[l, mapsto] \\
Z(I) := \{\underline{a} \in k^n \mid f(\underline{a}) = 0, \forall f \in I\} & I. \arrow[l, mapsto]
\end{tikzcd}$$

Definition 21.3.4. An **algebraic subset** is a subset of k^n of the form $Z(I)$ for some ideal $I \subseteq k[x_1, \dots, x_n]$.

Also note that $Z(I) = Z(\sqrt{I})$ because if $f^n(\underline{a}) = 0$, we must have $f(\underline{a}) = 0$. So it is natural to only consider radical ideals.

Theorem 21.3.5 (Hilbert's Nullstellensatz, strong form)**.** *There is a one-to-one correspondence between*

$$\begin{tikzcd}
\{\textit{Algebraic subsets of } k^n\} \arrow[r, leftrightarrow] & \{\textit{radical ideals of } k[x_1, \dots, x_n]\} \\
Z \arrow[r, mapsto] & I(Z) \\
Z(I) & I. \arrow[l, mapsto]
\end{tikzcd}$$

Remark 21.3.6. Algebraic subsets are considered "good spaces" in algebraic geometry: they are defined by polynomial equations. We can also talk about "polynomial functions" on an algebraic subset Z, defined by $\mathcal{O}(Z) := k[x_1, \dots, x_n]/I(Z)$. This makes sense because any function in $I(Z)$ vanishes on Z; so if two functions are differed by some functions in $I(Z)$, they would define the same function on Z.

Also, the maximal ideals of $\mathcal{O}(Z)$ are precisely of the form $\mathfrak{m}/I$ for some maximal ideal of $k[x_1, \dots, x_n]$, which by weak form of Hilbert Nullstellensatz is the same as $\mathfrak{m}_{\underline{a}}/I$ for some $\underline{a} \in k^n$. The condition $I \subseteq \mathfrak{m}_{\underline{a}}$ exactly implies that all functions in I vanishes at $\underline{a}$. So, in conclusion, the maximal ideals of $\mathcal{O}(Z)$ are exactly in one-to-one correspondence with the points in Z.

Extended reading after Lecture 21

21.4. **Construction of algebraic closure.** We give the proof of Theorem 21.1.5(1), namely, any field F is contained in an algebraically closed field K.

(Follow the textbook by M. Artin) Consider the following ring

$$R = F\big[x_f;\ \text{for each monic polynomial } f(x) \in F[x]\big].$$

In other words, this is to adjoin a free variable for every such polynomial. Put

$$I := \big(f(x_f);\ f \text{ monic polynomial}\big).$$

We claim that $I \neq (1)$. Indeed, if there exist $g_1(\underline{x}), \dots, g_r(\underline{x}) \in R$ such that

$$g_1(\underline{x}) f_1(x_{f_1}) + \cdots g_r(\underline{x}) f_r(x_{f_r}) = 1.$$

These g_i's only involve variables $x_1, \dots, x_m$, where each $x_i = x_{f_i}$ for $i = 1, \dots, r$. So we have

$$g_1(x_1, \dots, x_m) f_1(x_1) + \cdots + g_r(x_1, \dots, x_m) f_r(x_r) = 1. \tag{21.4.0.1}$$

Take a finite extension F' of F where each of $f_i(x)$ (with $i = 1, \dots, r$) has a root α_i. Evaluate (21.4.0.1) at $x_1 = \alpha_1$, ..., $x_r = \alpha_r$, $x_{r+1} = \cdots = x_m = 0$. This gives $0 = 1$. Contradiction!

By Zorn's lemma, there exists a maximal ideal $\mathfrak{m}_1$ of R containing I. Take $K_1 := R/\mathfrak{m}_1$. Then each polynomial in F has one zero in K_1. Continue this with F replaced by K_1, we get $K_2 = K_1[\cdots, x_h, \dots]/\mathfrak{m}_2$. Each polynomial in K_1 has one zero in K_2. Continue this way....

Define $K = \bigcup_{n \geq 1} K_n$. We claim that K is algebraically closed. This is because for any $f(x) = x^n + \cdots \in K[x]$, it belongs to some $K_m[x]$ for some m; then it splits in $K_{m+n}[x]$.

This K is an algebraically closed field containing F.

22. Noether normalization and Hilbert Nullstellensatz

Today, all rings are commutative.

22.1. **Integral ring extension.** Recall from the field extensions, we have proved that for a field extension K/F,

$$K/F \text{ is a finite extension } \iff K/F \text{ is finitely generated and algebraic.}$$

We develop the corresponding theory for rings.

Definition 22.1.1. Let $A \subseteq B$ be a subring. An element $x \in B$ is called **integral** over A if it satisfies a *monic* equation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$$

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

We point out that, since the polynomial ring over a general ring is no longer a PID, we do not have the notion of "minimal polynomial" here.

Proposition 22.1.2. *The following are equivalent.*

(1) $x \in B$ *is integral over* A*;*
(2) $A[x]$ *(= ring of all elements in* B *that can be expressed by a polynomial in* x *with coefficients in* A*) is a finitely generated* A*-module;*
(3) $A[x]$ *is contained in a subring* C *of* B *such that* C *is a finitely generated* A*-module.*

Proof. (1) ⇒ (2). Assume that x satisfies $x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$ for $a_0, \dots, a_{n-1} \in A$. So each x^{n+r} for $r \in \mathbb{Z}_{\geq 0}$ may be replaced by $-a_{n-1}x^{n+r-1} - \cdots - a_0x^r$. From this, we see that $A[x]$ is generated by $1, x, \dots, x^{n-1}$ as an A-module.

(2) ⇒ (3) Take $C = A[x]$.

(3) ⇒ (1) Assume that C is generated by $e_1, \dots, e_n$ as an A-module (not necessarily a basis; so there might be relations). We may write each xe_j (for $j = 1, \dots, n$) as an A-linear combination of this set of generators, i.e.

$$xe_j = a_{1j}e_1 + a_{2j}e_2 + \cdots a_{nj}e_n \quad \text{for} \quad a_{1j}, \dots, a_{nj} \in A.$$

(There might be more than one way to write xe_j; we take any such expression.) Writing this collectively, we have

$$(e_1, \dots, e_n)x = (e_1, \dots, e_n)\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

$$(e_1, \dots, e_n)\begin{pmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & x - a_{nn} \end{pmatrix} = 0.$$

Write S for the matrix on the right. By Cayley–Hamilton theorem, $\det(S)$ kills all elements $e_1, \dots, e_n$. But 1 is a linear combination of $e_1, \dots, e_n$. So $\det(S) = 0$; this shows that x is integral over A. □

Corollary 22.1.3. *Let $x_1, \dots, x_n$ be elements of B, each integral over A. Then $A[x_1, \dots, x_n]$ is a finitely generated A-module.*

Proof. For each i, assume that $x_i^{m_i} + a_{i,m_i-1}x_i^{m_i-1} + \cdots + a_{i,0} = 0$ for some $m_i \in \mathbb{N}$ and $a_{i,j} \in A$. Then $A[x_1, \dots, x_m]$ is generated as an A-module by monomials $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ with each $\alpha_i \in \{0, \dots, m_i - 1\}$. In particular $A[x_1, \dots, x_n]$ is a finitely generated A-module. □

Corollary 22.1.4. *The set C of elements of B which are integral over A is a subring of B containing A.*

Proof. Given $x, y \in C$, the previous corollary implies that $A[x, y]$ is a finitely generated A-module. By Proposition 22.1.2, $x \pm y, xy \in A[x, y]$ are integral over A; so $x \pm y, xy \in C$. Thus, C is a subring of B. □

Definition 22.1.5. This C in Corollary 22.1.4 is called the **integral closure** of A in B.

(1) If $C = A$, we say that A is **integrally closed** in B.
(2) If $C = B$, we say that B is **integral over** A.

Corollary 22.1.6. *If $A \subseteq B \subseteq C$ are rings and if B is integral over A and C is integral over B, then C is integral over A.*

Proof. (Compare with the proof of Theorem 14.4.14.) Let $x \in C$, the integrality implies that $x^n + b_{n-1}x^{n-1} + \cdots + b_0 = 0$ with $b_0, \dots, b_{n-1} \in B$. Consider the subring $B' = A[b_0, \dots, b_{n-1}] \subseteq B$. See the following diagram.

$$\begin{tikzcd}
C \arrow[d, dash] & \supseteq & B'[x] \arrow[d, dash] \\
B \arrow[d, dash] & \supseteq & B' = A[b_1, \dots, b_n] \\
A \arrow[urr, dash] & &
\end{tikzcd}$$

This B' is a finitely generated A-module as each $b_0, \dots, b_{n-1}$ is integral over A. Then $B'[x]$ is a finitely generated A-module, and hence x is integral over A. □

Corollary 22.1.7. *Let $A \subseteq B$ be rings and let C be the integral closure of A in B. Then C is integrally closed in B.*

Proof. If $x \in C$ is integral over B, then x is integral over A by the previous corollary. Thus $x \in C$. □

22.2. **Noether normalization.**

Definition 22.2.1. Let k be a field, a **finitely generated k-algebra** is a quotient $R = k[x_1, \dots, x_n]/I$ for some $n \in \mathbb{N}$ and some ideal I.

More generally, for A a ring, a **finitely generated A-algebra** is a quotient of $A[x_1, \dots, x_n]$.

Theorem 22.2.2 (Noether normalization)**.** *There exists some $r \leq n$ and an injective homomorphism*

$$\varphi : k[\underline{y}] = k[y_1, \dots, y_r] \hookrightarrow R$$

such that R is integral over $k[\underline{y}]$ (when we view $k[\underline{y}]$ as a subring of R).

Proof by Nagata. We prove the theorem by induction on n. Suppose that the theorem was proved when the ring R is generated by $n-1$ elements. Now if R is generated by n elements $x_1, \ldots, x_n$, i.e. $R = k[x_1, \ldots, x_n]/I$. If $I = (0)$, take $r = n$ and $y_i = x_i$ for $i = 1, \ldots, n$; we are done.

Now assume that $I \neq (0)$. Take a nonzero polynomial $f(\underline{x}) \in I$. Take positive integers $r_2, \ldots, r_n$ and put

$$z_2 = x_2 - x_1^{r_2}, \quad z_3 = x_3 - x_1^{r_3}, \quad \ldots, \quad z_n = x_n - x_1^{r_n}.$$

Consider the isomorphism

$$\begin{tikzcd}
k[x_1, \ldots, x_n] \arrow[d, "\cong"] & \supseteq & I \arrow[d, "\cong"] & \ni & f(x_1, \ldots, x_n) \arrow[d, mapsto] \\
k[x_1, z_2, \ldots, z_n] & \supseteq & \tilde{I} & \ni & \tilde{f}(x_1, z_2, \ldots, z_n)
\end{tikzcd}$$

The vertical arrow takes f to $\tilde{f}$. We will assume that $0 \ll r_2 \ll r_3 \ll \cdots \ll r_n$. Then $\tilde{f}$ has a unique leading term in x_1, namely

$$\tilde{f} = a \cdot x_1^N + \text{lower degree terms}.$$

So $k[x_1, \ldots, x_n]/(\tilde{f})$ is integral over $k[z_2, \ldots, z_n]$.

Note that the natural map $k[z_2, \ldots, z_n] \to R = k[x_1, z_2, \ldots, z_n]/\tilde{I}$ has kernel $\tilde{I} \cap k[\underline{z}]$. We put $R' = k[z_2, \ldots, z_n]/(\tilde{I} \cap k[\underline{z}])$, then we have a natural injection $R' \hookrightarrow R$, and may view R' as a subring of R. Consider the following diagram.

$$\begin{tikzcd}
k[x_1, z_2, \ldots, z_n]/(\tilde{f}) \arrow[r, two heads] \arrow[d, dash, "\text{integral}"'] & k[x_1, z_2, \ldots, z_n]/\tilde{I} = R & \\
k[z_2, \ldots, z_n] \arrow[r, two heads] & k[z_2, \ldots, z_n]/(\tilde{I} \cap k[\underline{z}]) = R' \arrow[u, "\text{integral}"] & k[y_1, \ldots, y_r] \arrow[l, "\text{integral}"']
\end{tikzcd}$$

The left vertical arrow the integral extension we just proved; both rings naturally surject to the middle column; the middle vertical arrow is then integral. This is because given any element s of R, it can be lifted to an element of the quotient $\tilde{s} \in k[x_1, z_2, \ldots, z_n]/(\tilde{f})$, which then satisfies an equation $\tilde{s}^n + \tilde{a}_{n-1}\tilde{s}^{n-1} + \cdots + \tilde{a}_0 = 0$ for some $n \in \mathbb{N}$ and $\tilde{a}_0, \ldots, \tilde{a}_{n-1} \in k[z_2, \ldots, z_n]$. Taking the image of this equation in R' shows that s is integral over R'. By inductive hypothesis, R' is generated over k by $n-1$ variables; so there exist some embedding $k[y_1, \ldots, y_r] \hookrightarrow R'$ such that R' is integral over $k[y_1, \ldots, y_r]$. By transitivity of integrality, R is integral over $k[y_1, \ldots, y_r]$. This completes the inductive proof of the Noether normalization theorem. □

Remark 22.2.3. The meaning of Noether normalization may be interpreted as: any finitely generated k-algebra is integral over some free k-algebra.

22.3. **Weak Hilbert Nullstellensatz.** First recall the statement of weak form of Hilbert Nullstellensatz theorem.

Theorem 22.3.1. *Let k be an algebraically closed field. Every maximal ideal $\mathfrak{m}$ of $k[x_1, \ldots, x_n]$ is of the form $\mathfrak{m}_{\underline{a}} = (x_1 - a_1, \ldots, x_n - a_n)$ for some $\underline{a} = (a_1, \ldots, a_n) \in k^n$.*

We now discuss the case when k is not algebraically closed. We start with an example.

Example 22.3.2. In $\mathbb{R}[x]$, (x^2+1) is a maximal ideal. Factor $x^2+1=(x+i)(x-i)$; so it corresponds to two points $x=i$ and $x=-i$. But none of the points belong to $\mathbb{R}$. Yet note that these two points are conjugate.

In general, we get an map

$$\begin{tikzcd}
\mathcal{M} : (k^{\mathrm{alg}})^n \arrow[r] & \Big\{\text{maximal ideals of } k[x_1,\dots,x_n]\Big\} \\
\underline{a}=(a_1,\dots,a_n) \arrow[r, mapsto] & \mathfrak{m}_{\underline{a}} := \ker\Big(k[x_1,\dots,x_n] \xrightarrow{\mathrm{ev}_{\underline{a}}} k(a_1,\dots,a_n)\subseteq k^{\mathrm{alg}}\Big).
\end{tikzcd}$$

Theorem 22.3.3 (weak Nullstellensatz for general fields)**.** *All maximal ideals of* $k[x_1,\dots,x_n]$ *arise this way.*

This theorem will be proved soon.

But $\mathcal{M}$ is not one-to-one. For each $\sigma\in\mathrm{Gal}(k^{\mathrm{alg}}/k)=\mathrm{Aut}(k^{\mathrm{alg}}/k)$, we get another point

$$\begin{tikzcd}
k[x_1,\dots,x_n] \arrow[r, "\mathrm{ev}_{\underline{a}}"] \arrow[rr, bend right, "\mathrm{ev}_{\sigma(\underline{a})}"'] & k^{\mathrm{alg}} \arrow[r, "\sigma"] & k^{\mathrm{alg}}.
\end{tikzcd}$$

Note that $\ker\mathrm{ev}_{\underline{a}}=\ker\mathrm{ev}_{\sigma(\underline{a})}$.

Theorem 22.3.4. *The map* $\mathcal{M}$ *induces a bijection*

$$\Big\{\mathrm{Gal}(k^{\mathrm{alg}}/k)\textit{-orbits of } (k^{\mathrm{alg}})^n\Big\} \longleftrightarrow \Big\{\textit{maximal ideals of } k[x_1,\dots,x_n]\Big\}.$$

Proof. We have seen that $\mathcal{M}$ is surjective and that $\mathfrak{m}_{\underline{a}}=\mathfrak{m}_{\sigma(\underline{a})}$.

Conversely, if $\ker\mathrm{ev}_{\underline{a}}=\ker\mathrm{ev}_{\underline{b}}=\mathfrak{m}$, then we have the following diagram.

$$\begin{tikzcd}
k[x_1,\dots,x_n] \arrow[r, two heads] \arrow[d, equal] & k[x_1,\dots,x_n]/\mathfrak{m} \arrow[r, phantom, "\cong"] \arrow[d, equal] & k(\underline{a}) \arrow[r, phantom, "\subseteq"] \arrow[d, "\eta", "\cong"'] & k^{\mathrm{alg}} \arrow[d, dashed, "\tilde{\eta}", "\cong"'] \\
k[x_1,\dots,x_n] \arrow[r, two heads] & k[x_1,\dots,x_n]/\mathfrak{m} \arrow[r, phantom, "\cong"] & k(\underline{b}) \arrow[r, phantom, "\subseteq"] & k^{\mathrm{alg}}.
\end{tikzcd}$$

Here the isomorphism $\eta : k(\underline{a}) \to k(\underline{b})$ is induced by that of the identifications with $k[x_1,\dots,x_n]/\mathfrak{m}$. This isomorphism extends to an isomorphism $\tilde{\eta} : k^{\mathrm{alg}}\cong k^{\mathrm{alg}}$. So $\tilde{\eta}(\underline{a})=\underline{b}$, i.e. $\underline{a}$ and $\underline{b}$ lie in the same orbit. □

We need a lemma before proving the Nullstellensatz.

Lemma 22.3.5. *Let* R *be a field and* $S\subseteq R$ *be a subring wuch that* R *is intgral over* S*. Then* S *is a field (and hence* R *is an algebraic extension of* S*).*

Proof. Clearly, S is an integral domain. It suffices to prove that $s\in S$ implies $s^{-1}\in S$.

Note $s^{-1}\in R$ is integral over S. So

$$s^{-n}+b_{n-1}s^{1-n}+\cdots+b_1s^{-1}+b_0=0.$$

$$s^{-1}=-b_{n-1}-b_{n-2}s-\cdots-b_0s^{n-1}\in S.$$

□

We now give the proof of Theorem 22.3.3.

Proof of weak Nullstellensatz Theorem 22.3.3. Let $\mathfrak{m}$ be a maximal ideal. Consider the following diagram.

$$\begin{tikzcd}
k[x_1, \dots, x_n] \arrow[r, two heads] & k[x_1, \dots, x_n]/\mathfrak{m} \\
 & k[y_1, \dots, y_r] \arrow[u, "\text{integral}"']
\end{tikzcd}$$

By Noether normalization, $k[x_1, \dots, x_n]/\mathfrak{m}$ is integral over some polynomial algebra $k[y_1, \dots, y_r]$. Yet $k[x_1, \dots, x_n]/\mathfrak{m}$ is a field; Lemma 22.3.5 implies that $k[y_1, \dots, y_r]$ is field; so $r = 0$.

Thus $k[x_1, \dots, x_n]/\mathfrak{m}$ is a finite extension of k; so it is a finite extension. Put $\ell := k[x_1, \dots, x_n]/\mathfrak{m}$. It embeds into k^{alg}. Write a_i for the image of x_i in the quotient ℓ. So we have

$$\begin{aligned} \mathfrak{m} \cong \big(k[x_1, \dots, x_n] &\longrightarrow \ell \subseteq k^{\mathrm{alg}}\big) \\ x_i &\longmapsto a_i. \end{aligned}$$

□

22.4. **Algebraic sets and Hilbert Nullstellensatz.**

Theorem 22.4.1 (Strong form of Nullstellensatz)**.** *Let k be an algebraically closed field. For an ideal $I \subseteq k[x_1, \dots, x_n]$, we have $I(Z(I)) = \sqrt{I}$.*

Proof. It is clear that $\sqrt{I} \subseteq I(Z(I))$: if $f \in \sqrt{I}$, then $f^n \in Z(I)$; so f^n vanishes on $Z(I)$ and thus f vanishes on $Z(I)$. So $f \in I(Z(I))$.

Conversely, we want to show that $I(Z(I)) \subseteq \sqrt{I}$, i.e. if $I = (f_1, \dots, f_m)$ and if $g \in k[x_1, \dots, x_n]$ satisfies

$$\forall\, \underline{a} \in k^n,\ f_1(\underline{a}) = \cdots = f_m(\underline{a}) = 0 \ \Rightarrow\ g(\underline{a}) = 0, \tag{22.4.1.1}$$

then there exist some $\ell \in \mathbb{N}$ such that $g^\ell \in (f_1, \dots, f_m)$. (Here, we secretly assumed that I is finitely generated; this is true, but we do not prove it here. The proof in fact does not depend on this finite generation.)

Now we add one more variable and consider the ideal

$$J = I \cdot k[x_1, \dots, x_n, x_{n+1}] + (1 - g \cdot x_{n+1}) \subseteq k[x_1, \dots, x_{n+1}].$$

Let us give this a bit more explanation: the condition (22.4.1.1) can be interpreted as

$$\big\{\underline{a} \in k^n \,\big|\, f_1(\underline{a}) = \cdots = f_m(\underline{a}) = 0,\ g(\underline{a}) \neq 0\big\} = \emptyset. \tag{22.4.1.2}$$

The condition $g(\underline{a}) \neq 0$ is the same as: there exists $a_{n+1} \in k$ such that $a_{n+1} \cdot g(\underline{a}) = 1$. So (22.4.1.2) is equivalent to saying that:

$$\big\{(a_1, \dots, a_{n+1}) \in k^{n+1} \,\big|\, f_1(\underline{a}) = \cdots = f_m(\underline{a}) = 0, a_{n+1} \cdot g(\underline{a}) = 1\big\} = \emptyset.$$

So it is expected that the ideal J is in fact the unit ideal.

Case 1: $J \neq (1)$. Then J is contained in a maximal ideal $\mathfrak{m} \subseteq k[x_1, \dots, x_{n+1}]$. By weak Nullstellensatz, $\mathfrak{m} = (x_1 - a_1, \dots, x_{n+1} - a_{n+1})$ for some $a_1, \dots, a_{n+1} \in k$. Under the map

$$\varphi : k[x_1, \dots, x_{n+1}] \twoheadrightarrow k[x_1, \dots, x_{n+1}]/\mathfrak{m} = k,$$

we have for any i,

$$0 = \varphi(f_i) = f_i(a_1, \dots, a_n) \text{ as } f_i \text{ belongs to } J.$$

This implies by (22.4.1.1) that $g(\underline{a}) = 0$. Yet as $1 - x_{n+1}g(\underline{x}) \in J$, we have

$$0 = \varphi(1 - x_{n+1}g(\underline{x})) = 1 - a_{n+1}g(\underline{a}).$$

This contradicts with $g(\underline{a}) = 0$.

Case 2: $J = (1)$. So there are polynomials $h_1, \dots, h_{m+1} \in k[x_1, \dots, x_{n+1}]$, such that

$$1 = h_1 f_1 + \cdots + h_m f_m + (1 - x_{n+1}g)h_{m+1} \quad \text{in } k[x_1, \dots, x_{n+1}].$$

In $k(x_1, \dots, x_n)$, we substitute $x_{n+1} = g^{-1}$ to get

$$1 = (h_1 f_1 + \cdots + h_m f_m)(x_1, \dots, x_n, g^{-1}).$$

Clearing the g in the denominator shows that

$$g^\ell = h_1^* f_1 + \cdots + h_m^* f_m$$

for some new polynomial h_i^*. This shows that $g \in \sqrt{I}$. □

Theorem 22.4.2 (Full Nullstellensatz)**.** *Assume that k is a algebraically closed field. There is a one-to-one bijection between*

$$\begin{tikzcd}
\big\{\textit{Algebraic subsets of } k^n\big\} \arrow[r, leftrightarrow] & \big\{\textit{radical ideals of } k[x_1, \dots, x_n]\big\} \\
Z \arrow[r, mapsto] & I(Z) \\
Z(I) & I \arrow[l, mapsto]
\end{tikzcd}$$

Moreover, we have the following properties.

(1) $I_1 \subseteq I_2 \Leftrightarrow Z(I_1) \supseteq Z(I_2)$.
(2) $Z(I_1 + I_2) = Z(I_1) \cap Z(I_2)$.
(3) $Z(I_1 \cap I_2) = Z(I_1) \cup Z(I_2)$.

Proof. We have just proved that, if I is radical, then $I(Z(I)) = \sqrt{I} = I$.

Conversely, for an algebraic set $Z = Z(J)$, we may first assume that J is radical because $Z(J) = Z(\sqrt{J})$. Now, we have

$$Z(I(Z)) = Z(I(Z(J)) = Z(J) = Z.$$

(1) and (2) are obvious.

(3) It is clear that $Z(I_1 \cap I_2) \supseteq Z(I_1) \cup Z(I_2)$. We need to show that $Z(I_1 \cap I_2) \subseteq Z(I_1) \cup Z(I_2)$. Suppose that $z \notin Z(I_1) \cup Z(I_2)$, then there exists $f_1 \in I_1$ and $f_2 \in I_2$ such that $f_1(z) \neq 0$ and $f_2(z) \neq 0$. Thus, $f := f_1 f_2 \in I_1 \cap I_2$ and $f(z) \neq 0$. It then follows that $z \notin Z(I_1 \cap I_2)$. So $Z(I_1 \cap I_2) \subseteq Z(I_1) \cup Z(I_2)$. □

Final note on page 163:

$$e^{\pi\sqrt{163}} = 262\,537\,412\,640\,768\,743.999\,999\,999\,999\,25\cdots \approx 640\,320^3 + 744.$$