Abstract algebra (a graduate course)

after P.A. Grillet [Gri07]

GTM 2007, vol. 242

Contents

Introduction

We will assume that the reader is already familiar with the basic notions of linear algebra such as the field of real and complex numbers, vector spaces, linear maps, kernel and image of a linear map. We have summarized the most important notions that will be needed for this course in the Appendix A.

Part I

First Semester

Chapter 1

Groups (群理论, 5 weeks)

1.1 Definition and Examples

Definition 1.1.1 (定义)**.** *A* group (群) *is a set (集合) G together with a binary operation $\cdot\colon G \times G \to G$ subject to the rules:*

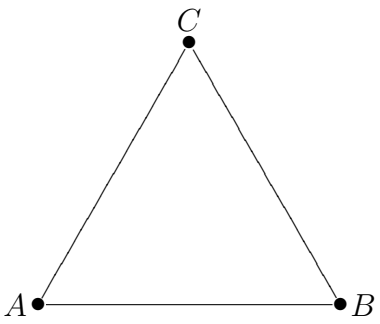
$$\forall a, b, c \in G\colon a(bc) = (ab)c, \tag{1.1}$$

$$\exists \mathrm{id} \in G\colon \forall a \in G\colon a \cdot \mathrm{id} = a = \mathrm{id} \cdot a, \tag{1.2}$$

$$\forall a \in G\colon \exists a^{-1} \in G\colon aa^{-1} = \mathrm{id} = a^{-1}a. \tag{1.3}$$

We define $(G : 1) := \operatorname{ord} G = |G|$ the cardinality of the group G. For finite groups this is the number of elements.

Example 1.1.2. 1.



Consider an equilateral triangle (i.e. all 3 sides have the same length) and its symmetry operations (对称操作). Beside the identity (身份), we have 3 reflections (反思, on a straight line through each corner) σ_A, σ_B, and σ_C, and 2 rotations (回转) τ (counter-clockwise) and τ^{-1} (clockwise).

In order to see that these are all symmetries, observe that a symmetry is uniquely determined by the image of the three corners A, B, C which can be permuted. The number of permutations is $3! = 6$ which correspond exactly to the above group elements. The group operation is composition of the symmetry operations. Note that in general the order of composition is

important. We make therefore the convention that in $f \circ g$ the operation g is applied first and then f. The multiplication table is therefore

id	τ	τ^{-1}	σ_A	σ_B	σ_C
τ	τ^{-1}	id	σ_C	σ_A	σ_B
τ^{-1}	id	τ	σ_B	σ_C	σ_A
σ_A	σ_B	σ_C	id	τ	τ^{-1}
σ_B	σ_C	σ_A	τ^{-1}	id	τ
σ_C	σ_A	σ_B	τ	τ^{-1}	id

To see that this operation is associative, note that for any family of maps (from a space back to itself, called *endomorphism*, 自同态) $(f \circ g)(\Delta) := f(g(\Delta))$ and therefore $((f \circ g) \circ h)(\Delta) = (f \circ g)(h(\Delta)) = f(g(h(\Delta))) = f((g \circ h)(\Delta)) = (f \circ (g \circ h))(\Delta)$ where Δ ranges over all objects that can be mapped. Therefore composition of endomorphisms is always associative.

We make two observations in this table. First, every row contains every element exactly once (also every column contains every element exactly once). Second, in the upper left 3×3-corner there are only the elements id, τ, and τ^{-1} – this is called a subgroup. Note that indeed multiplication is *not* commutative, e.g. $\sigma_A \circ \sigma_B = \tau \neq \tau^{-1} = \sigma_B \circ \sigma_A$. This group is denoted D_3 the symmetry group of the equilateral triangle. It can also be denoted S_3, because it permutes the 3 corners.

2. Consider the integers (整数) $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$ together with addition. The neutral element is obviously 0 (because $a+0 = a = 0+a$) and the inverse element to a is $-a$ (because $a+(-a) = 0 = (-a)+a$). Note that the group is infinite, but its operation is moreover *commutative* (交换), i.e. $b+a = a+b$. This group is denoted $(\mathbb{Z}, +)$.

3. Remainders of dividing integers modulo (模) $n > 0$ are $\{0, 1, 2, \dots, n-1\}$. a and b have the same remainder iff $n|(b-a)$. We thus arrange the integers in the classes (等价类) $[0] = \{0, \pm n, \pm 2n, \pm 3n, \dots\}$, $[1] = \{\dots, -2n+1, -n+1, 1, n+1, 2n+2, \dots\}$, $[2] = \{\dots, -2n+2, -n+2, 2, n+2, 2n+2, \dots\}$, ..., $[n-1] = \{\dots, -n-1, -1, n-1, 2n-1, \dots\}$. We write

$$a \equiv b \pmod{n} \Leftrightarrow a - b = kn \text{ for some } k \in \mathbb{Z}.$$

$a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$ then $a + c \equiv b + d \pmod{n}$, because $a - b = kn$ and $c - d = ln$ imply $(a + c) - (b + d) = (k + l)n$. With this we

can compute the following addition table

$[0]$	$[1]$	$[2]$	$[3]\ldots$	$[n-1]$
$[1]$	$[2]$	$[3]$	$[4]\ldots$	$[0]$
$[2]$	$[3]$	$[4]$	$[5]\ldots$	$[1]$
$\ldots$	$\ldots$			
$[n-1]$	$[0]$	$[1]$	$[2]\ldots$	$[n-2]$

This group is denoted as $(\mathbb{Z}/(n), +)$.

4. Let $p = 7$ and consider multiplication of the above classes modulo p. For $a \equiv b \pmod{p}$ and $c \equiv d \pmod{p}$ we also note that $ac \equiv bd \pmod{p}$, because $a - b = kp$ and $c - d = lp$ imply $ac - bd = (a-b)c + b(c-d) = kpc + blp = (kc + bl)p$. Moreover multiplication with 0 gives 0 which can therefore not be an element of the multiplicative group (has no inverse). The multiplication table now looks as follows

$[1]$	$[2]$	$[3]$	$[4]$	$[5]$	$[6]$
$[2]$	$[4]$	$[6]$	$[1]$	$[3]$	$[5]$
$[3]$	$[6]$	$[2]$	$[5]$	$[1]$	$[4]$
$[4]$	$[1]$	$[5]$	$[2]$	$[6]$	$[3]$
$[5]$	$[3]$	$[1]$	$[6]$	$[4]$	$[2]$
$[6]$	$[5]$	$[4]$	$[3]$	$[2]$	$[1]$

This group is denoted $(\mathbb{Z}/(p))^*$.

5. As an example of an associative (联想, and commutative) operation that does *not* form a group, consider the multiplication table modulo $n = 6$

$[1]$	$[2]$	$[3]$	$[4]$	$[5]$	$[0]$
$[2]$	$[4]$	$[0]$	$[2]$	$[4]$	$[0]$
$[3]$	$[0]$	$[3]$	$[0]$	$[3]$	$[0]$
$[4]$	$[2]$	$[0]$	$[4]$	$[2]$	$[0]$
$[5]$	$[4]$	$[3]$	$[2]$	$[1]$	$[0]$
$[0]$	$[0]$	$[0]$	$[0]$	$[0]$	$[0]$

These elements do not form a group, because e.g. $[2]$ has no inverse. However the elements $(\mathbb{Z}/(6))^* = \{[1], [5]\} \cong C_2$ do form a group with respect to multiplication.

Q: What causes the difference in the multiplication tables of $p = 7$ and $n = 6$?
a: p is a prime while n is not.

The most difficult part is to *check for associativity*. This can be done in a systematic way as proposed by Light: For every element $y \in G$ there is a table. The row headers (行头) are a copy the column (列) $\cdot y$ from the multiplication table. The row $y\cdot$ of the multiplication table labels the columns. The row of xy is copied to the according row in the table. If the column of yz in Light's table coincides with the column of $\cdot yz$ in the multiplication table, then $(xy)z = x(yz)$ for all $x \in G$. Correspondingly y passes Light's test if all columns are correct (i.e. $x(yz) = (xy)z$ for all $x, z \in G$). Finally the operation $\cdot$ is associative iff all elements pass Light's test.

id	τ	τ^{-1}	$\boldsymbol{\sigma_A}$	σ_B	σ_C
τ	τ^{-1}	id	$\boldsymbol{\sigma_C}$	σ_A	σ_B
τ^{-1}	id	τ	$\boldsymbol{\sigma_B}$	σ_C	σ_A
$\to \sigma_A$	σ_B	σ_C	**id**	τ	τ^{-1}
$\Rightarrow \sigma_B$	σ_C	σ_A	$\boldsymbol{\tau^{-1}}$	id	τ
σ_C	σ_A	σ_B	$\boldsymbol{\tau}$	τ^{-1}	id

$\to (\sigma_A)$	σ_B	σ_C	id	τ	τ^{-1}
$\boldsymbol{\sigma_C}$	.	.	τ	.	.
$\Rightarrow \boldsymbol{\sigma_B}$	σ_C	σ_A	$\tau^{-1} = \tau^{-1}$	id	τ
id	.	.	σ_A	.	.
$\boldsymbol{\tau^{-1}}$	.	.	σ_B	.	.
$\boldsymbol{\tau}$	.	.	σ_C	.	.

Table for $\sigma_A \in S_3$.

Lemma 1.1.3 (引理)**.** *Given a set X together with a binary operation $\cdot$. Let further X be generated (产生) by $S \subset X$ under the operation. If every element of S passes Light's test, then $\cdot$ is associative.* □

Definition 1.1.4. *Given a set X together with a binary operation $\cdot\colon X \times X \to X$, we say that $(X, \cdot)$ is a* semigroup *(半群) iff $\cdot$ is associative. If $\cdot$ has in addition a neutral element, we say that $(X, \cdot)$ is a* monoid *(幺半群).*

A (semi)-group is called commutative / abelian *(可交换的) if*

$$ba = ab \quad \forall a, b \in X.$$

Note that in a semi-group the inverse element is unique (独特), because $ea = a = ae$ and $e'a = a = ae'$ for all $a \in X$ imply $e = ee' = e'$. Also the inverse element is unique (if it exists), because $ai = e = ia$ and $aj = e = ja$ for some $a \in X$ imply $j = iaj = i$. A bit more interesting question is how much we can weaken the group axioms, i.e. do we really need that the inverse element is inverse from both sides?

Consider therefore the set of all sequences (序列) Seq $:= \mathbb{R}^\infty$ together with the operation $T : \text{Seq} \to \text{Seq} : (a_n) \mapsto (0, a_{n-1})$, i.e. the sequence is shifted to the right by inserting a leading 0. Obviously the operation $S : \text{Seq} \to \text{Seq} : (a_n) \mapsto (a_{n+1})$ that shifts sequences to the left (and drops the first element) fulfills $ST = \text{id}$, but on the other hand $TS : \text{Seq} \to \text{Seq} : (a_n) \mapsto (0, a_n)$ is not the identity, but sets the first element of the sequence to 0.

The example with the equilateral triangle has the following generalization:

Definition 1.1.5. *The dihedral group (二面体群) D_n is the symmetry group of the regular n-gon in the plane.*

Example 1.1.6. 3. $D_3 = \{\mathrm{id}, \tau, \tau^{-1}, \sigma_A, \sigma_B, \sigma_C\}$.

4. $D_4 = \{\mathrm{id}, \tau, \tau^2, \tau^{-1}, \sigma_A, \sigma_a, \sigma_B, \sigma_b\}$ where τ is a rotation by 90^o, σ_A the reflection on the diagonal AC, and σ_a the reflection on the perpendicular bisector of a (and of c).

n. $D_n = \{\mathrm{id}, \tau, \tau^2, \ldots, \tau^{n-1}, \sigma, \sigma\tau, \ldots, \sigma\tau^{n-1} : \tau^k\tau^l = \tau^{k+l}, \sigma^2 = \mathrm{id}, \tau^k\sigma = \sigma\tau^{-k}\}$

1.1.99 Exercises

Exercise 1.1.1. Given a semi-group $(X, \cdot)$ with a left-neutral element (i.e. $e \in X$ such that for all $a \in X$: $ea = a$) and left-inverses (for all $a \in X$ there is an $a_L \in X$ such that $a_L a = e$), show that $(X, \cdot)$ is a group.

What happens when we require a right-neutral and right-inverse elements?

Exercise 1.1.2. Let $(X, \cdot)$ be a semi-group and assume for every $a, b \in X$ the equations $ax = b$ and $ya = b$ have a solution. Show that $(X, \cdot)$ is a group.

Exercise 1.1.3. Let $(X, \cdot)$ be a finite semi-group. Assume that for every $a \in X$ the cancellation law (取消法则) holds, i.e. $ab = ac$ implies $b = c$ and $ba = ca$ implies $b = c$. Show that $(X, \cdot)$ is a group.

Given an example of an infinite semigroup where the cancellation law holds, but that is not a group.

Exercise 1.1.4. Describe the group of symmetries of the sine curve ($y = \sin x$ over the real numbers, 实数), i.e. list all its elements and write a multiplication table (compactly).

Exercise 1.1.5. Given a group $(G, \cdot)$. Show that $a^m a^n = a^{m+n}$ for all $a \in G$ and $m, n \in \mathbb{Z}$ where a^m has the usual meaning, i.e. $a^m = \underbrace{aa \cdots a}_{m \text{ times}}$ for $m > 0$, $a^0 = \mathrm{id}$ and $a^{-m} = (a^m)^{-1}$. Show moreover $(a^m)^n = a^{mn}$ for the same elements.

Exercise 1.1.6. Show that a finite group with an even number of elements contains an even number of elements x such that $x^{-1} = x$.

State and prove a similar statement for finite groups with an odd number of elements.

1.2 Subgroups (子群) and homomorphisms (同态)

Subgroups

Definition 1.2.1. *Given a group $(G, \circ)$, then a subset $H \subset G$ containing the neutral element* $\mathrm{id} \in H$ *and all inverse $h^{-1} \in H$ of its elements $h \in H$ that is closed under the operation $H \circ H \subset H$ is called a subgroup (子群).*

We denote $\langle g \rangle \subset G$ the subgroup generated by (所产生) $g \in G$ and more generally $\langle S \rangle \subset G$ the smallest subgroup containing the subset $S \subset G$.

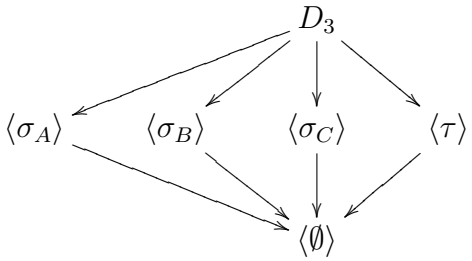
We denote $\operatorname{ord} g := (\langle g \rangle : 1)$ *the order of an element $g \in G$.*

We call a group $(G, \cdot)$ cyclic *(循环) iff there is an element $g \in G$ such that $G = \langle g \rangle$.*

Example 1.2.2. 0. The standard examples are $(\mathbb{Z}, +) \subset (\mathbb{Q}, +) \subset (\mathbb{R}, +) \subset (\mathbb{C}, +)$ as well as $\mathbb{Q}^* \subset \mathbb{R}^* \subset \mathbb{C}^*$.

1. Consider the group $(\mathbb{Z}, +)$. We can start from an element $n \in \mathbb{Z}$ to form the subgroup generated by n as $\langle n \rangle_{\mathbb{Z}} = n\mathbb{Z}$. Conversely given a finite number of elements $\{n_1, \ldots, n_k\}$ then the subgroup generated by them is the group $d\mathbb{Z}$ where $d = \gcd(n_1, \ldots, n_k)$ is the greatest common divisor. Therefore all subgroups of $\mathbb{Z}$ are cyclic.

2. Consider the group $D_3 = \{\mathrm{id}, \sigma_A, \sigma_B, \sigma_C, \tau, \tau^{-1}\}$. Its subgroups are $\{\mathrm{id}\} = \langle \emptyset \rangle$, $\{\mathrm{id}, \sigma_A\} = \langle \sigma_A \rangle$, $\langle \sigma_B \rangle$, $\langle \sigma_C \rangle$, $\langle \tau \rangle$, and D_3. These fit into the following scheme (子群方案)



Proposition 1.2.3 (命题)**.** *Given a group $(G, \cdot)$ then a non-empty subset $H \subset G$ is a subgroup iff for all $x, y \in H$ also $xy^{-1} \in H$.*

Proof. Since there is an $h \in H$, $\mathrm{id} = hh^{-1} \in H$. But then also $h^{-1} = \mathrm{id}h^{-1} \in H$. □

Proposition 1.2.4. *Given a* finite *(有限) group $(G, \circ)$, then a non-empty subset $H \subset G$ is a subgroup iff for all $x, y \in H$ also $xy \in H$.*

As the example of the natural numbers $\mathbb{N} \subset \mathbb{Z}$ shows, the assumption finite is important.

Proof. We want to use the previous proposition, but first need to show that for every $h \in H$ also $h^{-1} \in H$. We consider the subset $\langle h\rangle_+ := \{h^n : n \in \mathbb{N}_+\} \subset H$. Since G is finite, also H and thus $\langle h\rangle_+$ is finite. Therefore there are $m-1 > n \in \mathbb{N}_+$ such that $h^m = h^n$. Since there is $h^{-1} \in G$. we conclude that $\mathrm{id} = h^{m-n} \in H$. But then also $h^{-1} = h^{m-n-1} \in H$ which completes the proof. □

Note that in general the union (并集) of subgroups does not lead to a subgroup. However in the following special case it does.

Proposition 1.2.5. *Given a group $(G, \cdot)$ and an ascending chain of subgroups $H_0 \subset H_1 \subset H_2 \subset \cdots \subset G$, then their union $\bigcup_{n\geq 0} H_n$ is a subgroup of G.*

The interesting part is of course when the chain is infinite.

Proof. Denote the union by H_∞ and note first that $\mathrm{id} \in H_0$ and thus also in the union. Let $g \in H_\infty$. Then there is a (finite) number $m \in \mathbb{N}$ such that $g \in H_m$. But then also $g^{-1} \in H_m \subset H_\infty$, because H_m is a group. Let finally $g, h \in H_\infty$. Then there are (finite) numbers $m, n \in \mathbb{N}$ such that $g \in H_m$ and $h \in H_n$. Let $M = \max(m, n)$ be their maximum, then also $g, h \in H_M$ and thus their product is in H_M and thus in H_∞. □

In the homework you will prove that the intersection (交集) of subgroups always is a subgroup.

Definition 1.2.6. *Given a subgroup $H \subset G$ of a group, we define its left-cosets (左陪集) as $G/H := \{gH : g \in G\}$. The set of right-cosets (右～) is $H\backslash G := \{Hg : g \in G\}$.*

For abelian groups the two notions coincide, however in general they can be different.

Proposition 1.2.7 (Lagrange's theorem, 拉格朗日定理)**.** *Given a subgroup $H \subset G$ then its left-cosets (right-cosets) are disjoint. Moreover all cosets have the same cardinality as H. For G finite we have $\operatorname{ord} H | \operatorname{ord} G$ and in particular that the number of left-cosets equals the number of right-cosets. Also the order of every element $g \in G$ divides the group order $\operatorname{ord} g | \operatorname{ord} G$.*

Proof. The bijection $g \mapsto g^{-1}$ with $gh \mapsto h^{-1}g^{-1}$and thus $Hg \mapsto gH$ maps right-cosets bijectively onto left-cosets. It is therefore sufficient to prove the statements for left-cosets.

Given two cosets g_iH, $i = 1, 2$ whose intersection $g \in g_1H \cap g_2H$ is nonempty. Then there exist $h_i \in H$ such that $g = g_ih_i$. Therefore also $gh_i^{-1} = g_i$ and so $g_1H = gH = g_2H$.

For the second statement note that $gh_1 = g_1 = gh_2$ (with $h_{1/2} \in H$) implies by multiplication with g^{-1} from the left that $h_1 = h_2$. Therefore the map $H \to gH : h \mapsto gh$ is injective and surjective from H to the orbit gH for every $g \in G$.

In the last statement we just count $\operatorname{ord} G = |G/H| \operatorname{ord} H$. This completes the proof. □

We denote $[G : H] := |G/H|$ the *index* of H in G.

Remark 1.2.8. The relation $g \sim g'$ iff $gH = g'H$ is an equivalence relation (等价关系), i.e. $g \sim g$, $g \sim g' \Rightarrow g' \sim g$, and $g \sim g' \wedge g' \sim g'' \Rightarrow g \sim g''$ for all $g, g', g'' \in G$. Every equivalence relation breaks a set into disjoint equivalence classes.

Example 1.2.9. Let $G = D_3$ and $H = \langle \sigma_A \rangle$. The equivalence classes G/H are $\{H, \sigma_B H, \sigma_C H\}$ i.e. there are $3!/2 = 3$ equivalence classes. Note that the set G/H does not have any induced group structure, e.g. $H(\sigma_B H)$ has 4 elements.

Group homomorphisms

Definition 1.2.10. *Given two groups G and H. A group homomorphism (群同态) is a map $\phi\colon G \to H$ such that $\phi(g_1 g_2) = \phi(g_1)\phi(g_2)$.*

We denote $\operatorname{im}\phi = \{\phi(g) : g \in G\}$ *the* image *(图像) of ϕ and* $\ker\phi := \{g \in G : \phi(g) = \mathrm{id}\}$ *the* kernel *(内核) of ϕ.*

A group homomorphism $\phi\colon G \to H$ is called injective (monomorphism, 单同态) *iff* $\ker\phi = \{\mathrm{id}\}$*, it is called* surjective (epimorphism, 满同态) *iff* $\operatorname{im}\phi = H$*. It is called* bijective (isomorphism, 双同态) *iff it is injective and surjective. If there is a group isomorphism $\phi\colon G \to H$ we call G and H isomorphic, denoted as $G \cong H$.*

Note that the composition of group homomorphisms $\phi\colon H \to K$ and $\psi\colon G \to H$ is $\phi \circ \psi\colon G \to K : g \mapsto \phi(\psi(g))$ and is a group homomorphism, because $(\phi \circ \psi)(g_1 g_2) = \phi(\psi(g_1)\psi(g_2)) = (\phi \circ \psi)(g_1)(\phi \circ \psi)(g_2)$.

Also note that for a group homomorphism $\phi\colon G \to H$, $\phi(\mathrm{id}_G) = \mathrm{id}_H$, $\phi(g^{-1}) = (\phi(g))^{-1}$ and $\phi(g^n) = (\phi(g))^n$.

Example 1.2.11.

0. Given any two groups G and H and let $\mathrm{id}_H \in H$ denote its neutral element. Then $\phi_0\colon G \to H : g \mapsto \mathrm{id}_H$ is a (rather trivial) group homomorphism.

   Another example of a group homomorphism is the identity: $\mathrm{Id}_G\colon G \to G : g \mapsto g$.

1. A non-trivial example would be e.g. $e\colon (\mathbb{Z}, +) \to (\mathbb{R}, +)$ a monomorphism (also called an embedding, 包埋).

Proposition 1.2.12. *Given a subgroup $H \subset G$, then the image under a group homomorphism $\phi\colon G \to G'$ is also a subgroup $\phi(H) \subset G'$. Conversely, for $J \subset G'$ also $\phi^{-1}(J) := \{g \in G : \phi(g) \in J\}$ is a subgroup.*

Proof. In the first case note that $g_{1/2} \in \phi(H)$ imply $h_{1/2} \in H$ with $\phi(h_i) = g_i$. But then $h_1 h_2 \in H$ with $\phi(h_1 h_2) = \phi(h_1)\phi(h_2) = g_1 g_2$. Analogously for $g_1 \in \phi(H)$, $\phi(h_1^{-1}) = ((\phi(h_1))^{-1} = g_1^{-1} \in \phi(H)$ and $\mathrm{id}' = \phi(\mathrm{id}) \in \phi(H)$.

In the second case note that $g_{1/2} \in \phi^{-1}(J)$ implies $\phi(g_i) \in J$ and thus $\phi(g_1 g_2) = \phi(g_1)\phi(g_2) \in J$ which means $g_1 g_2 \in \phi^{-1}(J)$. Further $\phi(\mathrm{id}) = \mathrm{id}' \in J$ and therefore $\mathrm{id} \in \phi^{-1}(J)$. Finally $\phi(g^{-1}) = (\phi(g))^{-1} \in J$ and therefore $g^{-1} \in \phi^{-1}(J)$ for every $g \in \phi^{-1}(J)$. □

Corollary 1.2.13 (推论)**.** *Given a group homomorphism $\phi\colon G \to H$, then $\operatorname{im}\phi \subset H$ and $\ker\phi \subset G$ are subgroups. Moreover the left-cosets $G/\ker\phi$ coincide with the right-cosets $\ker\phi\backslash G$, i.e. $g\ker\phi = (\ker\phi)g$ for all $g \in G$.*

Proof. The first statement follows immediately from the last proposition, when you notice that $\operatorname{im}\phi = \phi(G)$ and $\ker\phi = \phi^{-1}(\mathrm{id})$ which are both trivial subgroups.

Then $g\ker\phi \subset \phi^{-1}(h)$ for $h = \phi(g)$, because $\phi(g\ker\phi) = \phi(g)\phi(\ker\phi) = h$. Conversely $\phi(g') = h = \phi(g)$, then $\phi(g^{-1}g') = \mathrm{id}_H$ and so $g^{-1}g' = g_0 \in \ker\phi$, i.e. $g' = gg_0 \in g\ker\phi$. Analogously $(\ker\phi)\cdot g = \phi^{-1}(h)$. □

Normal subgroups

Definition 1.2.14. *A subgroup $H \subset G$ is called* normal subgroup *(正规子群) if the left-cosets coincide with the right-cosets, i.e. for all $g \in G$, $gH = Hg$.*

Note that a subgroup $N \subset G$ is normal iff $gNg^{-1} \subset N$ for all $g \in G$.

Example 1.2.15. Consider the group $S_3 = D_3$ and the map $\mathrm{sgn}\colon S_3 \to \{\pm 1\} : \{\mathrm{id}, \tau, \tau^{-1}\} \to \{1\}, \{\sigma_A, \sigma_B, \sigma_C\} \to \{-1\}$. As we can check, this gives a group homomorphism, i.e. $\mathrm{sgn}(\alpha\beta) = \mathrm{sgn}(\alpha)\,\mathrm{sgn}(\beta)$ for all $\alpha, \beta \in S_3$. Its image is the whole group $H := \{\pm 1\}$. Its kernel is the subgroup $A_3 := \ker\mathrm{sgn} = \{\mathrm{id}, \tau, \tau^{-1}\}$ which is therefore a normal subgroup.

Proposition 1.2.16. *Given a normal subgroup $N \triangleleft G$, then the set of cosets inherits a group structure together with the surjective group homomorphism $\pi\colon G \to G/N : g \mapsto gN$ called* quotient map.

Proof. We want to define the group operation by representatives, i.e.

$$(gN)(hN) := (gh)N,$$

but for this we need to check that this is representative independent (代表性的独立). This is easy to see, because $Nh = hN$ by the preliminaries and thus $(gN)(hN) = gNhN = ghNN = ghN$, i.e. the coset multiplication is indeed the

element-wise multiplication of the cosets, but this is independent of the representatives g_i of the coset g_iN as well as for h_i and h_iN. Now it is also clear that π is a group homomorphism.

Finally surjectivity of the map π follows from the definition of G/N. □

Example 1.2.17. Given again the subgroup $A_3 \triangleleft S_3$, we have already proved that it is a normal subgroup. The two cosets are A_3 and $S_3 \setminus A_3 = \sigma_A A_3$. Therefore the group structure on S_3/A_3 is isomorphic to $(\{\pm 1\}, \cdot)$.

Corollary 1.2.18. *Given a normal subgroup $N \triangleleft G$, then the subgroups of G/N are in 1:1-correspondence with subgroups $\{S \subset G : N \subset S\}$.*

Proof. Let us denote $\pi\colon G \to G/N$ the projection. Given a subgroup $S \subset G$, then it projects onto a subgroup $\pi(S) \subset G/N$. Conversely, given $S' \subset G/N$ then it comes from a subgroup $\pi^{-1}(S') \subset G$ that contains $N = \pi^{-1}(\mathrm{id}')$. To see that this is the only subgroup containing N and projecting onto S', note that $\pi(g) \in S'$ is equivalent to $gN \in S$. But then all of gN must be contained in the subgroup $S \subset G$ that projects onto S'. □

Corollary 1.2.19. *Given a normal subgroup $N \triangleleft G$, then π and π^{-1} map inclusions to inclusions (i.e. $N \subset S_1 \subset S_2 \subset G$, then $\pi(S_1) \subset \pi(S_2)$ and the analogue for π^{-1}) and normal subgroups onto normal subgroups (i.e. $K \triangleleft G$ implies $\pi(K) \triangleleft G/N$).*

The proof of preservation of normality is left as an exercise.

In the search for normal subgroups we also obtain the following two notions:

Definition 1.2.20. *Given a group G, then we define*

1. *The center (中心) of G as* $\operatorname{cent} G := \{z \in G : \forall g \in G : zg = gz\}$,
2. *Given an abelian subgroup $H \subset G$ the centralizer (中心化子)* $\operatorname{cent}_G H := \{g \in G : \forall h \in H : gh = hg\}$,
3. *Given a subgroup $H \subset G$ its normalizer (正规化子)* $N_G(H) := \{g \in G : gH = Hg\}$.

Example 1.2.21. Consider the group D_4. Its center is trivial, i.e. $\{\mathrm{id}\}$ as the multiplication table shows. Consider further its subgroup $H = \langle \sigma_A \rangle$. It is not normal, because $\sigma_a H \neq H\sigma_a$, but it is abelian. Its centralizer is $\operatorname{cent}_{D_4} H = \langle \sigma_A, \sigma_C, \tau^2 \rangle$ and its normalizer is also $N_{S_4}(H) = \langle \sigma_A, \sigma_C, \tau^2 \rangle$.

Remark 1.2.22. Given a group G, then its center is a normal abelian subgroup. Given an abelian subgroup $H \subset G$, then its centralizer is the biggest subgroup of elements commuting with all elements of H. (In particular $H, \operatorname{cent}(G) \subset \operatorname{cent}_G(H)$.) Given any subgroup $H \subset G$, then its normalizer is the biggest subgroup in which H is a normal subgroup.

Definition 1.2.23. *Given a group G, then the* endomorphisms *(自同态)* $\mathrm{End}(G)$ *are the homomorphism of G into itself. The* automorphisms *(自同构)* $\mathrm{Aut}(G)$ *are the isomorphisms of G onto itself.*

Note that the endomorphisms form a monoid while the automorphisms form a group.

Example 1.2.24. Consider $C_3 = \mathbb{Z}/(3) = \langle [1] \rangle$. Obviously any endomorphism is specified by the image of $[1]$. The endomorphisms are therefore $\{[0], [1], [2]\}$ with the operation $\cdot$ (multiplication), the identity $\mathrm{Id}_{C_3} = [1]$. The automorphisms are those endomorphisms that are invertible, i.e. $\mathrm{Aut}(C_3) = \{[1], [2]\} = \mathrm{End}(C_3)^*$ under the same operation.

Summary: The study of groups means the study of group structure together with the subgroup structure, endo-, iso-, and other homomorphisms (to and from groups).

1.2.99 Exercises

Exercise 1.2.1. Given a group G and a family of subgroups $\{S_\alpha \subset G : \alpha \in A\}$. Show that

a. The intersection $\bigcap_{\alpha \in A} S_\alpha$ is a subgroup;

b. if all $S_\alpha \triangleleft G$ are normal, then the intersection is also normal.

Exercise 1.2.2. Let $G = D_n$ and $H = \{\mathrm{id}, \sigma\}$. Show that for $n \geq 3$ the partition into left-cosets is different from the partition into right-cosets.

Exercise 1.2.3.

a. Give a group together with two subgroups whose union is not a subgroup.

b. Given a group G together with two subgroups $S_{1/2} \subset G$. Show that $S_1 \cup S_2$ is a subgroup iff $S_1 \subset S_2$ or $S_2 \subset S_1$.

Exercise 1.2.4. Show that every group of prime order is simple (i.e. has only trivial subgroups) and cyclic.

Exercise* 1.2.5. Given a finite group G. A subgroup $M \subset G$ is called maximal iff it is different from G ($M \neq G$) and there is no subgroup properly in between ($M \subsetneq S \subsetneq G$). Show that every subgroup $H \subsetneq G$ that does not coincide with G is contained in a maximal subgroup.

Exercise 1.2.6. Given a group G together with two subgroups $H, K \subset G$. Show that the intersection of a left-coset of H with a left-coset of K is either empty or a left-coset of $H \cap K$.

Exercise 1.2.7. Given a group isomorphism $\phi\colon G \to H$ show that its inverse $\phi^{-1}\colon H \to G : \phi(g) \mapsto g$ is also a group homomorphism.

Exercise 1.2.8. Given a group homomorphism $\phi\colon G \to H$.

a. Assume that $N \triangleleft H$ is a normal subgroup. Show that $\phi^{-1}(N) \triangleleft G$ is a normal subgroup.

b. Assume that ϕ is surjective and $N \triangleleft G$ is a normal subgroup of G. Show that $\phi(N) \triangleleft H$ is a normal subgroup.

c. Find an example of a group homomorphism and a normal subgroup such that the image of the normal subgroup is not normal.

Exercise 1.2.9. Show that D_4 contains subgroups $A \subset N \subset D_4$ such that $A \triangleleft N$ and $N \triangleleft D_4$, but not $A \triangleleft D_4$.

Exercise 1.2.10. Prove that every subgroup of index 2 is normal.

Exercise 1.2.11. Prove that the union of an increasing sequence of normal subgroups $N_1 \subset N_2 \subset N_3 \subset \ldots$, $N_i \triangleleft G$ of a group G is normal.

Exercise 1.2.12. a. Let G be a group generated by $X \subset G$. Prove that for two homomorphisms $\phi, \psi\colon G \to H$ into any group H, $\phi(x) = \psi(x)$ for all $x \in X$ is equivalent to $\phi = \psi$.

b. Find all endomorphisms of $V_4 := \langle (12)(34), (13)(24), (14)(23) \rangle \subset S_4$ (Klein's four group).

c. Find all automorphisms of V_4.

d. Find all endomorphisms and automorphisms for D_3.

1.3 Isomorphism theorems (双同态定理)

The quotient map $G \mapsto G/N$ of a group G by a normal subgroup $N \triangleleft G$ gives an example of a universal map (也泛性质) in the following way.

Proposition 1.3.1. *Let $N \triangleleft G$ be a normal subgroup. Then every homomorphism $\phi\colon G \to H$ whose kernel contains $N \subset \ker \phi$ factors uniquely through the quotient map $\pi\colon G \to G/N$, i.e. there is a unique group homomorphism $\bar{\phi}\colon G/N \to H$ such that $\phi = \bar{\phi} \circ \pi$.*

Proof. The idea is to define $\bar{\phi}(gN) := \phi(g)$, but we first need to check that this is representation independent. Let thus $gN = hN$ for some $g, h \in G$. But then $\phi(g) = \phi(gN) = \phi(hN) = \phi(h)$ and thus $\bar{\phi}$ is well defined. The homomorphism properties follow now from those of ϕ. □

Theorem 1.3.2 (First isomorphism theorem)**.** *Given a group homomorphism $\phi\colon G \to H$, then $G/\ker\phi \cong \operatorname{im}\phi$.*

Proof. The obvious candidate for the isomorphism is $\bar{\phi}\colon G/\ker\phi \to \operatorname{im}\phi : g\ker\phi \mapsto \phi(g)$. First let us check that $\bar{\phi}$ is well defined. Let thus $N := \ker\phi$ and $gN = hN$ for some $g, h \in G$. But then $\phi(g) = \phi(gN) = \phi(hN) = \phi(h)$ and thus the image is the same. Second, note that $gN \in \ker\bar{\phi}$ implies $\bar{\phi}(gN) = \phi(g) = \mathrm{id}' \in H$ and thus $g \in N$. Therefore $\bar{\phi}$ is injective. Finally note that every $h \in \operatorname{im}\phi$ has a $g \in G$ with $\phi(g) = h$. But then $\bar{\phi}(gN) = h$ and thus $\bar{\phi}$ is also surjective and therefore bijective, i.e. an isomorphism. □

Example 1.3.3. Let $g \in G$ be an element, then $\langle g\rangle$ is cyclic. Consider the map $\phi\colon \mathbb{Z} \to G : n \mapsto g^n$.

If $m > n \in \mathbb{N}$ with $g^m = g^n$, then $g^{m-n} = \mathrm{id}$. Let now N be the smallest such difference. Then $g^m = \mathrm{id}$ for every $N|m$ and thus $\bar{\phi}\colon \mathbb{Z}/(N) \to \langle g\rangle$ is well-defined, maps $1 \mapsto g$ and thus also surjective. Moreover $\ker\phi = \mathbb{Z}/(N)$ and thus $\bar{\phi}$ is also injective, thus an isomorphism.

If there is no $m > n \in \mathbb{N}$ with $g^m = g^n$, then all the g^m are disjoint, moreover they are also disjoint from g^{-m} for any $m \in \mathbb{N}$. Thus $\phi\colon \mathbb{Z} \to \langle g\rangle$ is a homomorphism, surjective, and also injective. Therefore $\mathbb{Z} \cong \langle g\rangle$.

In particular every two cyclic groups of order n are isomorphic. We denote by C_n the cyclic group of order n.

Proposition 1.3.4. *For a cyclic group of order n there is for every divisor $d|n$ a unique subgroup of order d.*

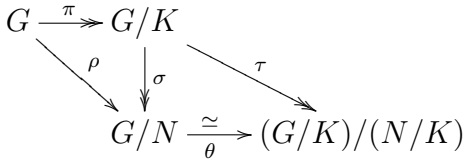
Proof. Let $C_n = \langle g\rangle$ be a cyclic group and g of order n. Define $S_d := \langle g^{n/d}\rangle$. Clearly S_d is a cyclic group which moreover has d elements, because for $h_0 := g^{n/d}$, $h_0^{d'} = \mathrm{id}$ with $0 \le d' < d$ would imply that h has order less than d and thus g order less than n which contradicts the assumption.

Conversely it is also possible to define a set $S'_d := \{h \in C_n : h^d = \mathrm{id}\}$, i.e. all those elements for which d is an exponent. Since $\langle g\rangle = C_n$, for every $h \in S'_d$ there is an $m \in \mathbb{N}$ such that $h = g^m$. Now $h^d = \mathrm{id}$ implies $dm \equiv 0 \pmod n$, i.e. $dm = kn$ for some $k \in \mathbb{N}$. But then $m = k\frac{n}{d}$ and thus $h = h_0^k$, i.e. $h \in S_d$ and thus $S'_d \subset S_d$. The other inclusion is obvious, and thus $S_d = S'_d$, i.e. both definitions coincide and the subgroup S_d is unique (i.e. independent of $g \in G$ as long as $\langle g\rangle = G$). □

Theorem 1.3.5 (Second isomorphism theorem)**.** *Let G be a group and $N, K \triangleleft G$ be normal subgroups. If $K \subset N$, then $K \triangleleft N$, $N/K \triangleleft G/K$ and*

$$(G/K)/(N/K) \cong G/N.$$

idea: $K \subset N \triangleleft G$ leads to

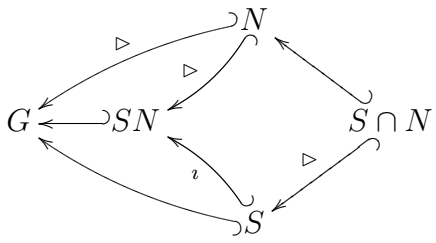


Proof. Let $\pi\colon G \to G/K$ and $\rho\colon G \to G/N$ be the quotient maps. We show that there is a unique isomorphism $\theta\colon G/N \to (G/K)/(N/K)$ such that $\theta \circ \rho = \tau \circ \pi$ where we also need to show that there is a morphism $\tau\colon G/K \to (G/K)/(N/K)$.

First note that ρ factors through π, because $K \subset N$, i.e. there is some homomorphism $\sigma\colon G/K \to G/N : gK \mapsto gN$ such that $\rho = \sigma \circ \pi$. Since ρ is surjective, so is σ. We show that $\ker \sigma = N/K$. First note that $K \triangleleft N$, because $K \triangleleft G$. For $n \in N$ we have $\sigma(nK) = nN = N$. Conversely if $\sigma(gK) = N$, then $gN = N$ and thus $g \in N$. This shows $\ker \sigma = N/K$. Therefore in particular $N/K \triangleleft G/K$. Now the first isomorphism theorem yields an isomorphism $\theta\colon G/N \xrightarrow{\sim} (G/K)/(N/K)$ such that $\theta \circ \sigma = \tau$. Then $\theta \circ \rho = \tau \circ \pi$. Since ρ is surjective, θ is unique with this property. This completes the proof. □

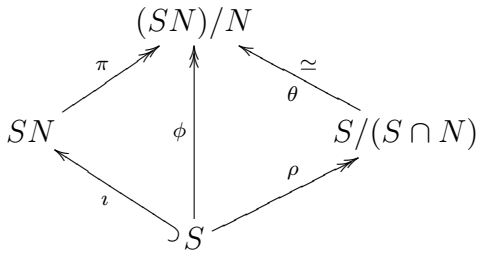
Theorem 1.3.6 (Third isomorphism theorem)**.** *Given a group G together with a subgroup $S \subset G$ and a normal subgroup $N \triangleleft G$, then $SN \subset G$ is a subgroup, $S \cap N \triangleleft S$ is a normal subgroup of S, and $S/(S \cap N) \cong (SN)/N$.*

The subgroup inclusion pattern is:



i.e. the arrows to the down-left indicate normal subgroups, the others are just subgroups, and the groups to the right (along an arrow) are subgroups of the

groups to the left. The required maps are



Proof. First note that $SN \subset G$ is indeed a group, because N is normal. Moreover $N \lhd (SN)$. We will show that there is a unique isomorphism $\theta\colon S/(S \cap N) \to (SN)/N$ such that $\theta \circ \rho = \pi \circ \imath$ where $\pi\colon SN \to (SN)/N$ is the projection, $\imath\colon S \hookrightarrow SN$ the inclusion, and there is a surjective homomorphism $\rho\colon S \to S/(S \cap N)$.

Let $\phi\colon S \to (SN)/N : s \mapsto sN$, i.e. $\phi = \pi \circ \imath$ and ϕ is surjective. Moreover $\phi(g) = N$ iff $g \in N$, i.e. $\ker \phi = S \cap N$ which is therefore normal in S. Therefore ρ is just the canonical projection (and in particular surjective). Again by the first isomorphism theorem $(SN)/N = \operatorname{im} \phi \cong S/\ker \phi = S/(S \cap N)$, i.e. there is an isomorphism $\theta\colon S/(S \cap N) \xrightarrow{\simeq} (SN)/N$. Moreover the isomorphism constructed in the proof of the theorem fulfills $\theta \circ \rho = \phi = \pi \circ \imath$ and is thus unique, because ρ is surjective. This completes the proof. □

The last isomorphism theorem implies in particular that the intersection of two normal subgroups of finite index has finite index.

1.3.99 Exercises

Exercise 1.3.1. Let $\phi\colon A \to B$ and $\psi\colon A \to C$ be group homomorphisms. Prove the following: If ψ is surjective, then ϕ factors through ψ if and only if $\ker \psi \subset \ker \phi$. In this case ϕ factors uniquely through ψ.

Exercise 1.3.2. Show that the identity homomorphism $\mathrm{Id}\colon 2\mathbb{Z} \xrightarrow{\sim} 2\mathbb{Z}$ does not factor through the inclusion homomorphism $\imath\colon 2\mathbb{Z} \hookrightarrow \mathbb{Z}$ even though $\ker \imath \subset \ker \mathrm{Id}$.

Hint: Opposite to the situation in Exercise 1.3.1, $\imath$ is not surjective.

Exercise 1.3.3. Let $\phi\colon A \to C$ and $\psi\colon B \to C$ be group homomorphisms. Prove the following: If ψ is injective, then ϕ factors through ψ if and only if $\operatorname{im} \phi \subset \operatorname{im} \psi$. In this case ϕ factors uniquely through ψ.

Exercise 1.3.4. Show that every subgroup of a cyclic group is cyclic.

Exercise 1.3.5. a. Show that the additive group $\mathbb{R}/\mathbb{Z}$ is isomorphic to the multiplicative group of all complex numbers $\mathbb{C}$ of modulus 1.

b. Show that the additive group $\mathbb{Q}/\mathbb{Z}$ is isomorphic to the group of all complex roots of unity (i.e. all complex numbers $z \neq 0$ such that $\langle z \rangle$ is finite in $\mathbb{C}^*$).

c. Show that the complex n-th roots of unity $\Omega_n := \{z \in \mathbb{C} : z^n = 1\}$ form a cyclic group (w.r.t. multiplication).

Exercise 1.3.6. Consider the group $D_4 := \langle \sigma, \tau : \sigma^2 = \mathrm{id} = \tau^4, \sigma\tau\sigma = \tau^{-1} \rangle$

a. Find the order of every element in D_4,

b. Show that for every $d|(D_4 : 1)$ there is a subgroup $S \subset D_4$ of order d.

Exercise 1.3.7. a. Let G be a finite group and $S, T \subset G$ any subgroups. Show that $|ST| = |S|\,|T|/|S \cap T|$.

b. Find a group G together with subgroups $S, T \subset G$ such that $ST \subset G$ is not a group.

Exercise 1.3.8. Let G be a finite group, $N \triangleleft G$ a normal subgroup and $H \subset G$ any subgroup such that $|N|$ and $(G : N)$ are relatively prime. Show that $H \subset N$ iff $|H|$ divides $|N|$.

Hint: Consider $HN \subset G$.

1.4 Free groups (自由群), free products (自由积), and presentations (群的展示)

Example 1.4.1. Consider the group $\mathbb{Z}/(4)$ it can be generated by the element $[1]$ (as well as $[3]$), because $[1] + [1] = [2], [1] + [2] = [3], [1] + [3] = [0]$. Moreover the elements fulfill the relations $a + (b + c) = (a + b) + c$, $[0] + a = a = a + [0]$, $b + a = a + b$, $4a := a + a + a + a = 0$ and infinitely many more.

Definition 1.4.2. *A presentation of a group is a set S of generators together with a set R of relations between them. Notation $G = \langle s \in S : R \rangle$.*

Lemma 1.4.3 (引理)**.** *Given a finite set S of generators there is a group $F(S)$ that is generated by S and for every map $\phi\colon S \to G$ into a group G there is a unique group homomorphism $\tilde{\phi}\colon F(S) \to G$.*

Example 1.4.4. 1. These groups are called *free groups* and the simplest example of a free group is $F_1 = \mathbb{Z}$ generated by 1. This is a cyclic group.

2. The free group on the generators S is (isomorphic to) the set of all (finite) canceled words in $S \cup \bar{S}$. A word $abc \ldots z$ is called canceled if there are no adjacent $a\bar{a}$ or $\bar{a}a$ for any $a \in S$. The group operation is concatenation together with canceling, i.e. $a\bar{a} \mapsto \epsilon$, $\bar{a}a \mapsto \epsilon$ for all $a \in S$.

Theorem 1.4.5. *Given a set S of generators and a set of relations (代数关系) R in elements of S, then there is a group generated by S that fulfills only the relations $\langle R\rangle_S$.*

Sketch of the proof. The idea is to start from the free group $F(S)$ and to impose the relations R. In order to do that we consider the normal subgroup $\langle R\rangle_S \triangleleft F(S)$ that is generated by R, i.e. $\langle R\rangle_S := \bigcap_{R\subset N\triangleleft F(S)} N$. The quotient is clearly a group generated by the images of S. □

Example 1.4.6.

1. The cyclic groups $\mathbb{Z}/(n) \cong \langle 1 : n\cdot 1 = 0\rangle$.
2. The dihedral group is $D_n := \langle \sigma, \tau : \sigma^2 = \mathrm{id} = \tau^n, \sigma\tau\sigma = \tau^{-1}\rangle$.
3. The *quaternions (四元数)* are defined as $\mathbb{H} := \mathbb{R}(\mathsf{i},\mathsf{j},\mathsf{k})$ where the unit quaternions i, j, and k multiply as $\mathsf{i}^2 = -1 = \mathsf{j}^2 = \mathsf{k}^2$ and $\mathsf{ij} = \mathsf{k} = -\mathsf{ji}$, $\mathsf{jk} = \mathsf{i} = -\mathsf{kj}$, $\mathsf{ki} = \mathsf{j} = -\mathsf{ik}$. These units form a group $Q = \{\pm 1, \pm\mathsf{i}, \pm\mathsf{j}, \pm\mathsf{k}\}$ (where in the homework you check the associativity). Obviously Q is generated by i and j, also $\mathsf{i}^4 = 1$ and $\mathsf{i}^2 = -1 = \mathsf{j}^2$. But we also have the relation $\mathsf{jij}^{-1} = \mathsf{i}^{-1}$. We want to show that these three generate all relations in Q, i.e. that $Q \cong \langle \mathsf{i},\mathsf{j} : \mathsf{i}^4 = \mathrm{id}, \mathsf{i}^2 = \mathsf{j}^2, \mathsf{jij}^{-1} = \mathsf{i}^{-1}\rangle$. Denote the second group by Q' and note that every element in Q' can be written as a finite sequence of is and js. Combining adjacent is to i^m and js to j^n, we see that $0 \le m, n \le 3$. Moreover the last relation ($\mathsf{ji} = \mathsf{i}^{-1}\mathsf{j}$) permits us to move every i to the left of all j. Therefore we are left with at most 16 elements. But we also see that we can replace j^2 by i^2 and thus j^3 by $\mathsf{i}^2\mathsf{j}$. Therefore we are left with 8 elements which are exactly the elements of Q and thus $Q' = Q$.

Corollary 1.4.7. *If G is a group generated by a subset $S \subset G$, then there is a unique surjective homomorphism $\phi\colon F(S) \to G$ that is the identity on S.* □

Corollary 1.4.8 (Free product)**.** *Given two groups G and H – where we assume $G \cap H = \{\mathrm{id}\}$ – there is a unique group $G * H$ that has $G \cup H$ as generators and fulfills only the relations $\langle R_G, R_H\rangle_{G\cup H}$.* □

Example 1.4.9. Given two free groups F_m and F_n then their free product $F_m * F_n \cong F_{m+n}$ is another free group.

Remark 1.4.10. A bit more difficult to show is that a subgroup of a free group is again a free group (possibly in 0 generators).

Remark 1.4.11. Beside the free product it is also possible to define the *amalgamation (共合积)*. Let thus $G \cap H = S$ be a joint subgroup. Denote $\gamma\colon G \hookrightarrow G * H$ and $\theta\colon H \hookrightarrow G * H$ the inclusions into the free product. Then $G *_S H =$

$G*H/\langle\gamma(s)\theta(s^{-1}) : s \in S\rangle_{G*H}$, i.e. the amalgamation of G and H over S is the quotient of the free product by the normal subgroup spanned by the anti-diagonal embedding of the intersection. It can be shown for $\bar{\gamma}\colon G \to G*_S H$ and $\bar{\theta}\colon H \to G*_S H$ that $G *_S H$ is generated by $\bar{\gamma}(G) \cup \bar{\theta}(H)$ and $\operatorname{im}\bar{\gamma} \cap \operatorname{im}\bar{\theta} = \bar{\gamma}(S) = \bar{\theta}(S)$.

Analogously to the free product, the amalgamated product fulfills the *universality property* that for every pair of group homomorphisms $\phi_G\colon G \to U$ and $\phi_H\colon H \to U$ such that $S = G \cap H$ and $\phi_G(S) = \phi_H(S)$, there is a unique $\tilde{\phi}\colon G *_S H \to U$ such that $\phi_G = \tilde{\phi} \circ \gamma$ and $\phi_H = \tilde{\phi} \circ \theta$.

1.4.99 Exercises

Exercise 1.4.1. Given a group G, the conjugates of an element $x \in G$ are $C_x := \{gxg^{-1} : g \in G\}$. Given a subset $S \subset G$, there exists a smallest normal subgroup $N \triangleleft G$ that contains $S \subset N$. Show that N consists of all products of elements in $C_{S \cup S^{-1}}$.

Exercise 1.4.2.

a. List (compactly) all elements of the group $\langle a, b : a^2 = \mathrm{id} = b^2\rangle$. Give a compact multiplication table of the group.

b. List all elements of the group $\langle a, b : a^2 = \mathrm{id} = b^2 = (ab)^3\rangle$ and give their multiplication table. Which known group is it isomorphic to?

Remark 1.4.12. Note that $\langle a : a^2 = \mathrm{id}\rangle \cong C_2 \cong \langle b : b^2 = \mathrm{id}\rangle$ are cyclic groups of order 2. Since the group in part a only fulfills the two relations $a^2 = \mathrm{id} = b^2$, it is (isomorphic to) the free product $C_2 * C_2$.

Exercise 1.4.3. The multiplication of the unit quaternions $\mathsf{i}^2 = -1 = \mathsf{j}^2 = \mathsf{k}^2$, $\mathsf{ij} = \mathsf{k} = -\mathsf{ji}$, $\mathsf{jk} = \mathsf{i} = -\mathsf{kj}$, $\mathsf{ki} = \mathsf{j} = -\mathsf{ik}$ together with $\mathbb{R}$-linearity implies for $a, b, c, d, a', b', c', d' \in \mathbb{R}$,

$$\begin{aligned}(a + b\mathsf{i} + c\mathsf{j} + d\mathsf{k})(a' + b'\mathsf{i} + c'\mathsf{j} + d'\mathsf{k}) = \\ = (aa' - bb' - cc' - dd') + (ab' + ba' + cd' - dc')\mathsf{i} \\ + (ac' + ca' + db' - bd')\mathsf{j} + (ad' + da' + bc' - cb')\mathsf{k}.\end{aligned}$$

a. Show that the multiplication is associative.

b. Let $\overline{a + b\mathsf{i} + c\mathsf{j} + d\mathsf{k}} := a - b\mathsf{i} - c\mathsf{j} - d\mathsf{k}$ and $|z|^2 := z\bar{z}$ for every quaternion $z \in \mathbb{H}$. Show that $|z_1 z_2| = |z_1|\,|z_2|$ for every pair of quaternions $z_{1/2} \in \mathbb{H}$.

c. Conclude that $\mathbb{H}^* := \mathbb{H} \setminus \{0\}$ is a group under multiplication. (What is the inverse? Therefore $\mathbb{H}$ is called a division algebra.)

Remark 1.4.13. It is possible to consider $\bar{z}z$ as a positive definite sesquilinear-form, namely $\langle z_1, z_2\rangle := \bar{z}_1 z_2$ which maps $\langle .,.\rangle \colon \mathbb{H}\times\mathbb{H}\to\mathbb{H}$, has

$$\langle z, w_1+\lambda w_2\rangle = \langle z, w_1\rangle + \lambda\langle z, w_2\rangle$$

for all $z_i, w_i \in \mathbb{H}$ and $\lambda\in\mathbb{R}$,

$$\begin{aligned}\langle z_2, z_1\rangle &= (a_2a_1+b_2b_1+c_2c_1+d_2d_1)+(a_2b_1-b_2a_1-c_2d_1+d_2c_1)\mathsf{i}\\ &\quad +(a_2c_1-c_2a_1-d_2b_1+b_2d_1)\mathsf{j}+(a_2d_1-d_2a_1-b_2c_1+c_2b_1)\mathsf{k}\\ &= \overline{\langle z_1, z_2\rangle},\\ |z|^2 := \langle z, z\rangle &= a^2+b^2+c^2+d^2\geq 0 \quad \text{and ``}=0\text{'' iff } z=0.\end{aligned}$$

I.e. $\langle .,.\rangle$ is indeed positive definite and sesquilinear.

1.5 Direct products (真积)

A much simpler construction is the following.

Definition 1.5.1. *Let $\{G_\alpha : \alpha\in A\}$ be groups. Their direct product $\prod_{\alpha\in A} G_\alpha$ is the set*

$$\left\{(g\colon A\to\bigcup_{\alpha\in A} G_\alpha) : g_\alpha\in G_\alpha\right\}$$

under component-wise operation.

Proposition 1.5.2. *The direct product of two groups G_i, $i=1,2$ is the group generated by $G_1\cup G_2$ under the relations $\langle R_1, R_2, ab=ba : a\in G_1, b\in G_2\rangle_{G_1\cup G_2}$.*

Proof. Let $\bigoplus$ denote the construction in the statement. We want to show that $\theta\colon \bigoplus\to\prod : g_1g_2\mapsto(g_1,g_2)$ where $g_i\in G_i$ is a group isomorphism. First note that θ is well-defined, because the elements of G_1 and G_2 commute and we can thus sort all factors in G_1 into the first component and the others into the second component. For $g,h\in\oplus$ we have $\theta(gh)=\theta(g_1g_2h_1h_2)=\theta(g_1h_1g_2h_2)=(g_1h_1,g_2h_2)=(g_1,g_2)(h_1,h_2)=\theta(g)\theta(h)$ and thus θ is a group homomorphism. Note also that θ is surjective. Since $G_1\cap G_2=\{\mathrm{id}\}$, we see that θ is also injective. Therefore θ is the required isomorphism. □

Example 1.5.3. Given the two groups C_2 and C_4 then their direct product is the group $C_2\times C_4=\{(a,b) : a\in C_2, b\in C_4\}$ with component-wise operation.

Remark 1.5.4. Note however that for infinite products, e.g. $\mathbb{R}^\infty$, the two notions differ. Namely while $\bigoplus_{n\in\mathbb{N}}\mathbb{R}$ is the abelian group of all finite sequences (which is separable, 可分空间), $\prod_{n\in\mathbb{N}}\mathbb{R}$ is the abelian group of all sequences (which is not separable, even if we restrict to bounded sequences).

Q: Given elements $g \in G$ and $h \in H$ of finite order. What is the order of (g, h) in $G \times H$?
a: $\operatorname{ord}(g, h) = \operatorname{lcm}(\operatorname{ord} g, \operatorname{ord} h)$.

The question we want to answer next is when can a group be written as direct product of subgroups.

Proposition 1.5.5. *Given a group G together with two subgroups $G_i \subset G$, $i = 1, 2$, then $G \cong G_1 \times G_2$ iff both $G_i \triangleleft G$ are normal, $G_1 \cap G_2 = \{\mathrm{id}\}$, and $G_1 G_2 = G$.*

Proof. First consider the group $G' := G_1 \times G_2$. The two projections $\pi_i \colon G' \to G_i : (g_1, g_2) \mapsto g_i$ show that each $G_i \triangleleft G'$. Moreover $G' = G_1 G_2$ and $G_1 \cap G_2 = \{(\mathrm{id}, \mathrm{id})\} \subset G'$. Therefore the conditions are necessary.

Assume now that $G_i \triangleleft G$ are both normal, $G_1 \cap G_2 = \{\mathrm{id}\}$, and $G_1 G_2 = G$. We consider the homomorphism $\theta \colon G_1 \times G_2 \to G : (g_1, g_2) \mapsto g_1 g_2$. If $g_1 g_2 = g = g_1' g_2'$, then $g_1^{-1} g_1' = g_2 g_2'^{-1} \in G_1 \cap G_2 = \{\mathrm{id}\}$, i.e. $g_1 = g_1'$ and $g_2 = g_2'$. Also for every $g \in G$ we have $g_i \in G_i$ such that $g_1 g_2 = g$. Therefore θ is a bijection (of sets). It remains to show that θ is a homomorphism. If we could show that the elements $g_i \in G_i$ commute, the proof would be complete. Note that $g_1(g_2 g_1^{-1} g_2^{-1}) \in G_1$, because $G_1 \triangleleft G$ is normal. Conversely $(g_1 g_2 g_1^{-1}) g_2^{-1} \in G_2$. Therefore $g_1 g_2 g_1^{-1} g_2^{-1} = \mathrm{id}$ and thus $g_1 g_2 = g_2 g_1$ which completes the proof. □

Example 1.5.6. The easiest examples arise from abelian groups, because for those all subgroups are normal. Consider, e.g. $C_6 \cong \mathbb{Z}/(6)$ the additive group of integers modulo 6. Two subgroups are $C_2 = \{[0], [3]\}$ and $C_3 = \{[0], [2], [4]\}$ which obviously intersect in $C_2 \cap C_3 = \{[0]\}$ and $C_2 C_3 = C_6$. Therefore $C_6 \cong C_2 \times C_3$.

This generalizes to arbitrary (finitely generated) abelian groups as follows.

Theorem 1.5.7 (Structure thm for abelian groups, 阿贝尔群的结构定理)**.** *Given a finitely generated abelian group A, then there are primes $p_i \in \mathbb{P}$ together with positive integers $n_i \in \mathbb{N}_+$ and $n_0 \in \mathbb{N}$ such that*

$$A \cong \mathbb{Z}^{n_0} \bigoplus_{i \geq 1} \mathbb{Z}/(p_i^{n_i}).$$

These pairs are unique up to order.

We call $\operatorname{Tor}(A) := \{a \in A : \operatorname{ord} a < \infty\}$ the *torsion part (挠子群)* of A and $\operatorname{rk} A := n_0$ the *rank (秩)* of $\mathbb{Z}^{n_0}$ (or also A if $\operatorname{Tor} A = 0$).

Proof. First note that the sum is finitely generated abelian. Conversely, let $\operatorname{Tor} A$ be the torsion part of A and $A_\infty := A/\operatorname{Tor} A$ the free abelian (自由交换子群) part. Obviously this decomposition is unique as well as both parts finitely generated (as subgroup and quotient of a finitely generated group, respectively).

Next consider an abelian p-group A, i.e. a finite group of order p^k for some prime $p \in \mathbb{P}$ and natural number $k \in \mathbb{N}$. Obviously every element of a p-group has order a power of p. Let $N(i) := |\{a \in A : (p^i)a = 0\}|$. Let n be the highest i with $N(i) > 0$, then there is an element $a \in A$ of order p^n and $A/\langle a\rangle$ has $N'(n) < N(n)$. By induction (and Proposition 1.5.5) we can thus write $A \cong \bigoplus_i \mathbb{Z}/(p^{n_i})$. The construction also shows that the decomposition is unique up to order, because $\langle g\rangle$ of order p^n has exactly $p^n - p^{n-1}$ elements of order p^n, $p^{n-1} - p^{n-2}$ elements of order p^{n-1}, ..., and 1 element of order 1, in addition $a \times b$ has order $\operatorname{lcm}(\operatorname{ord} a, \operatorname{ord} b)$ in the direct product $A \times B$; therefore A/C_{p^n} always kills the same amount of elements of order p^k regardless of the generator $g \in A$.

Let now A be abelian of order $n = p_1^{n_1} \dots p_k^{n_k}$ with distinct primes $p_i \in \mathbb{P}$. We construct the subgroups $S_{p_i} = \{a \in A : (p_i^{n_i}) \cdot a = 0\}$ and see that they have trivial intersection (only the neutral element 0) and are unique. Conversely for every $a \in A$, $\langle a\rangle$ is cyclic and for $\operatorname{ord}_A a = p_1^{n_1'} \cdots p_k^{n_k'}$ there are elements of order $p_i^{n_i'}$ in $\langle a\rangle$, namely $a^{(\operatorname{ord} a)/p_i^{n_i'}}$ and thus a is the (finite) product of some elements of the S_{p_i}. Therefore $A \cong \bigoplus_{i \geq 1} S_{p_i}$. Together with the previous consideration this completes the decomposition of the torsion group.

Let now A_∞ be a torsion-free abelian group generated by N elements $a_1, \dots, a_N$. Its decomposition is analogous to p-groups, i.e. we consider the first generator a_1, the induced subgroup $A_1 := \langle a_1\rangle \subset A_\infty$ and the quotient $\pi \colon A_\infty \to A_c := A_\infty/A_1$. If $A_T := \operatorname{Tor} A_c = 0$, i.e. the quotient is torsion free, then $A_\infty = A_1 \times A_c$, and $A_c = \langle \pi(a_1), \dots, \pi(a_N)\rangle$. But the first generator vanishes, because $\langle a_1\rangle = A_1$. Moreover we also drop all other vanishing generators. Therefore A_c is generated by $a_1', \dots, a_{N'}'$ with $0 \leq N' \leq N-1$, i.e. the procedure stops after finitely many steps. We are left with the problem that A_T may not be 0. Since we only divided out 1 generator of a torsion-free group, A_T must be cyclic. Let thus A_T^* be the generators of the torsion group. We can thus find another $\tilde{a}_1 \in \pi_1^{-1}(A_T^*)$ such that $\langle a_1\rangle \subset \langle \tilde{a}_1\rangle$ (because $\pi\tilde{a}_1 \in A_T^*$ generated A_T) and now $\tilde{\pi} \colon A_\infty \to A_\infty/\langle \tilde{a}_1\rangle \cong A_c/A_T$ is torsion-free. The construction depends on the choice of a_1, and $\tilde{a}_1$, but it ensures that the chosen $\tilde{a}_1, \dots, \tilde{a}_{n_0}$ are a $\mathbb{Z}$-linear independent set of generators of A_∞. Thus an argument similar to the uniqueness of dimension of a vector space shows that the number n_0 is independent of these choices.[1] □

The last part of the proof used some interesting interplay between abelian groups and the integers which in parts resembled properties of vector spaces (e.g. you can call a $\mathbb{Z}$-linear independent set of generators a $\mathbb{Z}$-base). The useful context in which to generalize these ideas are modules (a generalization of vector spaces)

[1] In particular we can express any other $\mathbb{Z}$-linear independent set of generators by a linear combination with the coefficients of an invertible matrix C with entries in $\mathbb{Z}$, i.e. $C \in \mathrm{GL}_{n_0}(\mathbb{Z}) = \{C \in \operatorname{End}(\mathbb{Z}^n) : \det C \in \{\pm 1\} = \mathbb{Z}^*\}$.

over (principal ideal) domains (of which $\mathbb{Z}$ is an example). Further details can be found in Chapter 5.

Example 1.5.8. Consider all abelian groups of order $16 = 2^4$. The structure theorem states that they are the direct product of cyclic subgroups of order 2^{n_i} such that $2^4 = 16 = \prod_i 2^{n_i}$, i.e. $4 = \sum_i n_i$. Obviously we can arrange the summands n_i in increasing order $n_1 \leq n_2 \leq n_3 \leq \dots$. We thus need the *partitions* of 4 into positive integers. These are (4), $(1,3)$, $(1,1,2)$, $(1,1,1,1)$, and $(2,2)$. These correspond to the 5 non-isomorphic groups C_{2^4}, $C_2 \times C_{2^3}$, $(C_2)^2 \times C_{2^2}$, $(C_2)^4$, and $(C_{2^2})^2$.

1.5.99 Exercises

Exercise 1.5.1 (Product and coproduct, 直积与对偶直积)**.** Given two groups G and H.

a. Show that their direct product $G \times H$ together with the canonical projections $\pi_G \colon G \times H \to G$ and $\pi_H \colon G \times H \to H$ is a product (in the sense of category theory), i.e. for every pair of homomorphisms $\rho_G \colon K \to G$ and $\rho_H \colon K \to H$ there is a unique morphism $\tilde{\rho} \colon K \to G \times H$ such that $\pi_G \circ \tilde{\rho} = \rho_G$ and $\pi_H \circ \tilde{\rho} = \rho_H$.

b. Show that $G \oplus H = G \times H$ together with the canonical embeddings $e_G \colon G \to G \oplus H$ and $e_H \colon H \to G \oplus H$ is a coproduct, i.e. for every pair of homomorphisms $\rho_G \colon G \to K$ and $\rho_H \colon H \to K$ with $\rho_G(g)\rho_H(h) = \rho_H(h)\rho_G(g)$ for all $g \in G$ and $h \in H$, there is a unique morphism $\tilde{\rho} \colon G \oplus H \to K$ such that $\rho_G = \tilde{\rho} \circ e_G$ and $\rho_H = \tilde{\rho} \circ e_H$.

c. Draw the mapping diagrams for the above constructions. You could draw every given map by a solid arrow and every map implied through your proof by a dashed/ dotted arrow.

Exercise 1.5.2. Find all abelian groups (up to isomorphism) of order

a. 35,

b. 36,

c. 360.

Exercise 1.5.3. Define Euler's ϕ-function as $\phi \colon \mathbb{N}_+ \to \mathbb{N}_+$ such that $\phi(n)$ is the cardinality of $\{k \in [1,n] : \gcd(k,n) = 1\}$, i.e. the numbers that are relatively prime to n. Show the following formula for ϕ:

$$\phi(n) = n \prod_{p|n} (1 - \tfrac{1}{p})$$

where $p \in \mathbb{P}$ runs over all prime divisors of n.

You can proceed as follows:

a. Show that the group C_p has exactly $p-1$ generators, i.e. there are $p-1$ numbers between 1 and p (inclusive) that are relatively prime to p;

b. Show that the group C_{p^n} has exactly $p^n - p^{n-1}$ elements of order p^n, i.e. $\phi(p^n) = p^n(1 - \frac{1}{p})$;

c. Show that the group C_{mn} where $\gcd(m,n) = 1$ has the generators $(C_m)^* \times (C_n)^*$ where $(C_m)^*$ are the generators of C_m. Conclude that $\phi(mn) = \phi(m)\phi(n)$ and thus the formula for ϕ.

Exercise 1.5.4. A group G is called *indecomposable* iff for every direct sum $G \cong A \oplus B$ either $A = 1$ or $B = 1$.

a. Prove that D_5 is indecomposable;

b. prove that D_4 is indecomposable;

c. prove that C_{p^k} is indecomposable when p is a prime and $k \in \mathbb{N}$.

1.6 The Krull–Schmidt theorem (克鲁尔–施密特定理)

Given a group we can ask whether it is isomorphic to any known group. In order to better see such isomorphisms we can ask whether we can decompose the group into (finitely many) minimal unique factors such that the whole group is the direct sum of the factors.

Given a decomposition as $\mathbb{Z} \cong \mathbb{Z}/(2) \times 2\mathbb{Z} \cong \mathbb{Z}/(2) \times \mathbb{Z}/(3) \times 6\mathbb{Z} \cong \mathbb{Z}/(2) \times \mathbb{Z}/(3) \times \mathbb{Z}/(5) \times 30\mathbb{Z} \cong \ldots$ it is clear that the answer in general in no. A positive answer can be given with the following notions:

Definition 1.6.1. *Given a group G, we define the following notions.*

1. *An ascending chain (升链) of subgroups is a sequence $\{\mathrm{id}\} \subset S_1 \subset S_2 \subset \cdots \subset G$, where all groups S_n are subgroups of G and all the later S_k with $k > n$.*

2. *A descending chain (降链) of subgroups is a sequence $G \supset S_1 \supset S_2 \supset \ldots$ such that every S_n is a subgroup of G as well as all the previous groups S_k with $k < n$.*

3. *An ascending / descending chain of subgroups is said to be finite if there is an n such that $S_n = S_{n+1} = S_{n+2} = \dots$ for all following subgroups.*

Example 1.6.2. 0. Given a finite group, then it has only finitely many subgroups and thus all ascending and descending chains of subgroups are finite.

1. Consider the group $(\mathbb{Z}, +)$, then we have an infinite chain of descending subgroups, namely $\mathbb{Z} \supset 2\mathbb{Z} \supset 6\mathbb{Z} \supset 30\mathbb{Z} \supset \dots$.

2. Consider the group $(\mathbb{Q}/\mathbb{Z}, +)$, i.e. the rational numbers modulo integers. These have an infinite chain of ascending subgroups as follows: $\{0\} \subset \{\frac{n}{2} : n \in \mathbb{Z}/(2)\} \subset \{\frac{n}{6} : n \in \mathbb{Z}/(6)\} \subset \{\frac{n}{30} : n \in \mathbb{Z}/(30)\} \subset \dots$.

It is clear that in the latter two cases we cannot expect any (finite or infinite) decomposition into indecomposable subgroups.

The following theorem, named after Wolfgang Krull[2] and Otto Schmidt[3] gives a positive answer to the classification problem.

Theorem 1.6.3 (Krull–Schmidt)**.** *Given a group G for which all ascending and all descending chains of normal subgroups are finite, then the group is the finite direct sum of indecomposable subgroups. In particular these factors are unique up to isomorphism including multiplicity.*

The proof is a bit lengthy and uses the idea of *normal endomorphisms (正常自同态)*, i.e. a group homomorphism $\phi\colon G \to G$ such that for all $g, h \in G$: $\phi(ghg^{-1}) = g\phi(h)g^{-1}$, i.e. ϕ commutes with all *inner automorphisms* $c_g\colon G \to G : h \mapsto ghg^{-1}$.

Example 1.6.4. Consider the finite groups $G = D_6 = \langle \sigma, \tau : \sigma^2 = \mathrm{id} = \tau^6, \sigma\tau\sigma = \tau^{-1} \rangle$ and $H = C_2 \oplus D_3$. They don't look isomorphic at first sight, but let us compute their Krull–Schmidt decomposition. In the first case we see that $D_3 \hookrightarrow D_6 : \sigma \mapsto \sigma, \tau \mapsto \tau^2$ is a subgroup of index 2 and thus normal. Also $C_2 \hookrightarrow D_6 : 1 \mapsto \tau^3$ commutes with the embedding of D_3. Thus $G \cong C_2 \oplus D_3$ and each of the factors are indecomposable. In the second case we obtain thus $G \cong H$.

1.6.99 Exercises

Exercise 1.6.1. Compute the Krull–Schmidt decomposition of the $D_n = \langle \sigma, \tau : \sigma^2 = \mathrm{id} = \tau^n, \sigma\tau\sigma = \tau^{-1} \rangle$ for $n \leq 8$.

[2] $*8/1899$ in Germany $\dagger 4/1971$

[3] $*9/1891$ in Russia $\dagger 9/1956$

Exercise 1.6.2. Compute the Krull–Schmidt decomposition of $\mathrm{GL}_2(\mathbb{F}_n)$ the automorphism group of the vector space $\mathbb{F}_n^2$ where $\mathbb{F}_n$ is the field with n elements (obviously $n \neq 6$ and some other cases).

a. For $n = 2$.

b. Count the number of elements in $\mathrm{GL}_2(\mathbb{F}_3)$.

c* For $n = 3$.

d. Count the number of elements in $\mathrm{GL}_2(\mathbb{F}_p)$ for $p \in \mathbb{P}$.

1.7 Group actions (群作用)

Remember the introducing example of symmetries of the equilateral triangle. It turned out that the group permutes the corners of the triangle. This concept can be generalized to arbitrary groups in the following way.

Definition 1.7.1. *Given a group G and a set (集合) X. An* action (作用) *of G on X is a binary operation $\mu\colon G \times X \to X : (g, x) \mapsto \mu(g)x$ such that $\mu(\mathrm{id}) = \mathrm{Id}_X$ and $\mu(gh) = \mu(g)\mu(h)$.*

We denote $Gx := \{gx : g \in G\}$ the orbit (轨道) *of G through $x \in X$ and $\mathrm{Stab}_G(x) := \{g \in G : \mu(g)x = x\}$ the* stabilizer (稳定子群) *of $x \in X$.*

We say that a group G acts transitively (可递的作用) *on a set X if for every pair of elements $x, y \in X$ in X there is a group element $g \in G$ such that $\mu(g)x = y$.*

Remark 1.7.2. By definition the group acts transitively on each orbit. Therefore intersecting orbits must coincide. Also transitivity from x to y is an equivalence relation (reflexivity $x \sim x$ via $\mathrm{id} \in G$, $y \sim x$ via g^{-1} if $x \sim y$ via $g \in G$, and transitivity $x \sim z$ via hg if $x \sim y$ via g and $y \sim z$ via $h \in G$.).

Remark 1.7.3. Correspondingly there is also the notion of a *right action (从右边的作用)* $\rho\colon X \times G \to X : (x, g) \mapsto x^g$. Note that this means in particular $\rho(gh) = \rho(h)\rho(g)$, i.e. $\rho\colon G \to S(X)$ is an anti-homomorphism. There is however a 1:1-correspondence between (left)- and right-actions via $\mu(g) := \rho(g^{-1})$.

Corollary 1.7.4 (of Lagrange's theorem, Stabilizer–Orbit Theorem)**.** *Given the action of a finite group G on a set X, then for every $x \in X$ we have* $\operatorname{ord} G = |Gx| \operatorname{ord} \mathrm{Stab}_G(x)$.

Proof. Note that the orbit of G through $x \in X$ is isomorphic to the left-coset $G/\mathrm{Stab}_G(x)$. □

As a particular example remember the conjugation action of a group G on itself: $c\colon G \to \mathrm{Aut}(G) : g \mapsto c_g$ with $c_g\colon G \to G : h \mapsto ghg^{-1}$. Clearly $c\colon G \to \mathrm{End}(G)$ is a monoid homomorphism, because $c_g(c_{g'}(h)) = gg'hg'^{-1}g^{-1} = (gg')h(gg')^{-1} = c_{gg'}(g)$, also $(c_{\mathrm{id}}\colon G \to G : h \mapsto \mathrm{id}h\mathrm{id}^{-1} = h) = \mathrm{Id}_G$ and thus $c_g^{-1} = c_{g^{-1}}$ is the inverse endomorphism and thus c_g indeed an automorphism for every $g \in G$.

We call the orbits of $x \in G$ under c the *conjugacy classes (共轭类)* of x. The conjugacy class of a central element $z \in \mathrm{cent}(G)$ is just $\{z\}$. The partition of G into conjugacy classes (its orbits under the conjugacy action) of a finite group gives a partition of the group order, i.e.

$$|G| = |\,\mathrm{cent}_G\,| + \sum_{|C|>1} |C|$$

where the sum runs over the non-trivial conjugacy classes of G. This is called the *class equation of the group (群的类方程).*

Example 1.7.5. 0. for an abelian group $G = \mathrm{cent}_G$ and thus the sum part is 0.

1. Remember the symmetry group of the square $D_4 = \langle \sigma, \tau : \sigma^2 = \mathrm{id} = \tau^4, \sigma\tau\sigma = \tau^{-1}\rangle$. Its center is $\mathrm{cent}_{D_4} = \{\mathrm{id}, \tau^2\}$ and the other two rotations form a conjugacy class$\{\tau, \tau^{-1}\}$.[4] The remaining reflections break into two conjugacy classes $\{\sigma\tau, \sigma\tau^{-1}\}$ and $\{\sigma, \sigma\tau^2\}$.[5] The class equation is therefore

$$|D_4| = 8 = 2 + 2 + 2 + 2.$$

1.7.99 Exercises

Exercise 1.7.1. Explain how the original statement of Lagrange's theorem "*When $x_1, \ldots, x_n$ are permuted in all possible ways, then the number of different values of $f(x_1, \ldots, x_n)$ is a divisor of $n!$.*" relates to orbits and stabilizers.

Exercise 1.7.2. Let G be a group and for $g \in G$ define the *inner automorphism (内自同构)* $c_g\colon G \to G : h \mapsto ghg^{-1}$.

a. Show that the inner automorphisms c_g form a subgroup $\mathrm{Inn}(G) \subset \mathrm{Aut}(G)$ isomorphic to $G/\,\mathrm{cent}(G)$.

b. Show that $\mathrm{Inn}(G) \triangleleft \mathrm{Aut}(G)$, i.e. a normal subgroup.

Hint: What is $(\phi \circ c_g)(h)$ for $g, h \in G$?

[4] because $\sigma\tau\sigma = \tau^{-1}$

[5] because $\tau\sigma\tau^{-1} = \sigma\tau^{-2} = \sigma\tau^2$ and $\sigma(\sigma\tau)\sigma = \tau\sigma = \sigma\tau^{-1}$

Exercise 1.7.3. Let $\mu\colon G \times X \to X$ be the action of a group G on a set X.

a. Let $x, y \in X$ be two points on the same orbit. Show that their stabilizers are conjugate, i.e. there is an element $g \in G$ such that $\mathrm{Stab}_G(x) = g\,\mathrm{Stab}_G(y)g^{-1}$.

b. Assume that $\mathrm{Stab}_G(x) \cong C_2$ and $\mathrm{Stab}_G(y) \cong C_3$. Can x and y be on the same orbit? (Justify your answer.)

Exercise 1.7.4. Show that in a finite group G of order n, an element of order k has at most n/k conjugates.

Exercise 1.7.5. Determine the class equation of the $D_n := \langle \sigma, \tau : \sigma^2 = \mathrm{id} = \tau^n, \sigma\tau\sigma^{-1} = \tau^{-1} \rangle$ where $n = 1, 2, \ldots$.

Exercise 1.7.6. Assume that $G/\operatorname{cent}(G)$ is cyclic. Prove that G is abelian.

Exercise 1.7.7. A *characteristic subgroup (特征子群)* $H \subset G$ is a subgroup that is invariant under all automorphisms, i.e. for all $\phi \in \mathrm{Aut}(G)$: $\phi(H) = H$. In particular characteristic subgroups are invariant under the inner automorphisms and therefore normal.

a. Show that the center $\operatorname{cent}(G)$ is a characteristic subgroup.

b. Prove that every characteristic subgroup $H \subset N$ of a normal subgroup $N \triangleleft G$ is normal $H \triangleleft G$ in G.

c. Assume that $N \triangleleft G$ is characteristic and $N \subset H \subset G$ with $H/N \subset G/N$ characteristic. Show that $H \subset G$ is characteristic.

1.8 Structure of symmetric groups (置換群)

Note that every group G acts on the space $X = G$ *faithful (忠实的作用)*[6] and transitively by left-multiplication $l(g)h = gh \in X$. We have therefore an isomorphism of every (finite) group to a subgroup of some (finite) permutation group (Cayley's theorem). It is therefore useful to study the structure of permutation groups.

Definition 1.8.1. *Given a (finite) set X. A* permutation (置換) *of X is a bijective map $\sigma\colon X \to X$.*

Lemma 1.8.2. *The permutations of a set X form a group, denoted $S(X)$ the symmetric group (置換群) on X.*

[6] G acts faithfully on X if $gx = x$ implies $g = \mathrm{id}$.

Proof. The composition of two permutations σ and τ of X is the map $\sigma \circ \tau \colon X \to X : x \mapsto \sigma(\tau(x))$ which is also a permutation. The neutral element is the identity map id_X. Given a permutation σ its inverse is the inverse map σ^{-1} which is also a permutation. □

The straight-forward way to write down permutations is as maps, e.g. $\sigma = \binom{123}{312}$. Note that this notation becomes unique if we require the top line (the preimages) to be sorted.

A more efficient notation is via the orbits (轨道), e.g. the permutation $\sigma = \binom{123456}{312456}$ has the two orbits (132) and (45) which are cycled in this order. We can therefore write $\sigma = (132)(45)$, because the orbits encode where every element of X is mapped (elements not listed are mapped to themselves). Note that it does not matter in which order you multiply disjoint orbits.

For the general multiplication of orbits, consider the following example: (1234)(24) – We first follow the element 4 which is first mapped to 2 and in the second permutation (the left one which is applied second) it is mapped to 3. Therefore the new orbits start with $(43\ldots)$. We then follow the element 3 which is mapped to 4. Therefore the first orbit reads (43). Next, we follow another element, say 2. This is mapped to 4 and then to 1. Therefore the second disjoint orbit starts as $(21\ldots)$. Now we follow the 1 and so on, to obtain the 2 orbits (43)(21). Since now all elements from the product are listed, this is the complete orbit notation of the product.

A *transposition (替换)* is an orbit of length 2.

Lemma 1.8.3. *Every permutation can be written as a product of transpositions.*

Proof. It is sufficient to show that for orbits. Orbits of length 2 are trivial. Any longer orbit can be numbered as $(12\ldots n)$ and thus decomposed as $(12)(23)\ldots(n-1\,n)$. □

Note that the decomposition of orbits into transpositions is not unique, e.g. $(12)(13)(12) = (23)$.

Lemma 1.8.4. *The parity (奇偶性) of the number of transpositions as a product of which an orbit can be written is opposite to the parity of the number of elements.*

An orbit is called *even* if it decomposes into an even number of transpositions.

Proposition 1.8.5. *Given an orbit* $(12\ldots n)$*. Its signum is 1 if it is even, it is* -1 *if it is odd.* $\mathrm{sgn}\colon S_n \to \pm 1$ *is a group homomorphism.*

Proof. Note that the notation of a permutation as disjoint orbits is unique. Therefore the number of odd orbits is well defined. Due to the last lemma this number adds up modulo 2 under multiplication. □

Definition 1.8.6. *Given a positive integer $n > 0$. The* alternating group (交错群) *A_n on n letters is defined as the kernel of the signum map* $\operatorname{sgn}\colon S_n \to \{\pm 1\}$.

Example 1.8.7.

2. The group S_2 consists of the identity and the flip (12). The alternating group is $A_1 = \{\mathrm{id}\}$.

3. The group S_3 consists of the 6 permutations $\{\mathrm{id}, (12), (13), (23), (123), (132)\}$. Its alternating group is $A_3 = \{\mathrm{id}, (123), (132)\} \cong C_3$.

4. The group S_4 has $4! = 24$ elements and its alternating group A_4 consists of the 12 elements $\{\mathrm{id}, (123), (132), (124), (142), (134), (143), (234), (243), (12)(34), (13)(24), (13)(24)\}$

5. The symmetric group S_5 consists of $5! = 120$ elements. Its alternating group A_5 consists of 60 elements (and is simple, 单群).

Conversely, we also have the following property:

Proposition 1.8.8. *A_n is generated by all 3-cycles.*

Proof. Note that every 3-cycle has odd length and is thus even (sic!). Conversely every element of A_n can be written as the product of an even number of transpositions. We are thus left to write the product of 2 transpositions $(ab)(cd)$ as products of 3-cycles.

If $\{a, b\} = \{c, d\}$, then $(ab)(cd) = \mathrm{id}$ and we thus need 0 3-cycles. If $\{a, b\}$ and $\{c, d\}$ are disjoint, we can write $(ab)(cd) = (ab)(bc)(cb)(cd) = (abc)(bcd)$. Finally, if $\{a, b\}$ and $\{c, d\}$ have one element in common, w.l.o.g. we denote it $b = c$ and write $(ab)(bd) = (abd)$. □

1.8.99 Exercises

Exercise 1.8.1.

a. Show that S_n is generated by (12), (23),..., $(n-1\,n)$.

b. Show that S_n is generated by (12) and $(12\ldots n)$.

Exercise 1.8.2.

a. Show that $S_4 \cong \langle a, b : a^4 = \mathrm{id} = b^2 = (ba)^3 \rangle$.

b. Show that $A_4 \cong \langle a, b : a^3 = \mathrm{id} = b^2, aba = ba^2b \rangle$.

Exercise 1.8.3. How many k-cycles are there in S_n?

Exercise 1.8.4. Consider the permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 7 & 5 & 6 & 4 & 2 & 8 & 3 & 1 \end{pmatrix}$

a. Write σ as product of disjoint orbits. Determine its signum and its order.

b. What is the order of the centralizer of σ in S_8? What is the order of the conjugacy class of σ?

Exercise 1.8.5. a. List all conjugacy classes of S_5 together with their orders.

b. List all conjugacy classes of A_5 together with their orders.

Warning: There are even cycles of S_5 that are conjugate in S_5 but not in A_5.

c. Conclude that A_5 has no normal subgroup beside 1 and A_5.

1.9 The Sylow theorems (西罗定理)

Consider the task of classifying all finite groups. Given the list of examples of finite groups we have obtained so far, you might first ask whether it is abelian or non-abelian. In the former case the Krull–Schmidt theorem allows you to decompose it into minimal cyclic groups whose order is a prime power. The order and multiplicity of these subgroups identifies the finite abelian group uniquely.

For finite non-abelian groups you can also decompose it into a direct product of indecomposable subgroups as soon as you have found normal subgroups. However the study of such groups is much harder. A helpful tool is the theorem due to Peter Ludwig Mejdell Sylow.[7] Before we can state that theorem, we need to introduce some notions.

Definition 1.9.1. *Given a prime $p \in \mathbb{P}$, then a finite group G is a p-group if $(G : 1) = p^K$ for some $K \in \mathbb{N}$.*

By Lagrange's Theorem every element in a p-group has order p^k for some integer $0 \leq k \leq K$ (Depending on the element).

Definition 1.9.2. *Given a finite group G of order n and a prime $p \in \mathbb{P}$. Then a p-Sylow subgroup is a p-subgroup of maximal order, i.e. there is no p-subgroup in which it is strictly contained.*

Theorem 1.9.3 (Sylow)**.** *Given a finite group G of order n and a prime $p \in \mathbb{P}$. Then the following statements hold about p-Sylow subgroups.*

1. *If $p^K | n$, but not $p^{K+1} | n$, then there exists at least one p-Sylow subgroup of order p^K.*

2. *All p-Sylow subgroups of G are conjugate and thus of order p^K.*

[7] *12/1832 in Norway, † 9/1918

3. *The number n_p of p-Sylow subgroups of G is congruent 1 modulo p and $n_p|m$ where $n = mp^K$.*

Proof. The proof covers the main part of this section. We follow a combinatorial proof sketched in [web, Sylow's theorems]. It consists of several lemmas.

Lemma 1.9.4. *Let $0 \leq k \leq K$ be an integer, then there is a subgroup of order p^k in G.*

Proof. Let $\Omega := \binom{G}{p^k}$ be the set of all subsets of G with p^k elements, and let G act on Ω by left-multiplication.[8] Since G is a group this is indeed an action. For any $\omega \in \Omega$ let $G_\omega := \mathrm{Stab}_G(\omega)$ be its stabilizer group. For any $a \in \omega \subset G$ the map $R_a \colon g \mapsto ga$ maps G_ω into ω injectively (the inverse map is $R_{a^{-1}}$). Therefore $p^k = |\omega| \geq |G_\omega|$.

On the other hand for $M := n/p^k$

$$|\Omega| = \binom{Mp^k}{p^k} = \prod_{j=0}^{p^k-1} \frac{Mp^k - j}{p^k - j} = M \prod_{j=1}^{p^k-1} \frac{Mp^k - j}{p^k - j}$$
$$= M \prod_{j=1}^{p^k-1} \frac{Mp^{k-\nu_p(j)} - j/p^{\nu_p(j)}}{p^{k-\nu_p(j)} - j/p^{\nu_p(j)}}$$

where the p-adic valuation $\nu_p(j) = k' \in \mathbb{N}$ such that $p^{k'}|j$ but $p^{k'+1} \nmid j$. In particular $\nu_p(j) \leq k$ for integers $1 \leq j \leq p^k - 1$. Therefore none of the factors in the last $\prod$ contains any p. So $\nu_p(|\Omega|) = \nu_p(M)\nu_p\left(\prod \dots\right) = \nu_p(M)$. Let $R \subset \Omega$ contain one representative per orbit, thus $|R| = |\Omega/\!\sim|$. Moreover

$$|\Omega| = \sum_{\omega \in R} |G\omega|.$$

Let $r := K - k$ and $s := \nu_p(|G\omega|) \leq \nu_p(|\Omega|) = r$. By the definition of ν_p, p does not divide $M' := |G\omega|/p^s \in \mathbb{N}_+$, which must therefore divide M. By Corollary 1.7.4 we have $|G_\omega| = |G|/|G\omega| = p^{k+r-s}M/M'$ and thus $p^k \,\big|\, |G_\omega|$, i.e. $p^k \leq |G_\omega|$. Together we obtain $|G_\omega| = p^k$, i.e. $G_\omega \subset G$ is the desired subgroup. □

This proves the first statement.

Next we want to show a property of p-group actions:

Lemma 1.9.5. *Let G be a finite p-group with an action on a finite set X. Let X_0 denote the set of fixed points. Then $|X| \equiv |X_0| \pmod p$.*

[8] If we could find a set $\omega \in \Omega$ that is a group, then we would be done. However this is rather hard, so instead we let G act on Ω and do the job for us.

Proof. Write X as disjoint union of its orbits under G. Any element $x \in X \setminus X_0$ will lie in an orbit of order $|G|/|\operatorname{Stab}_G(x)|$ which is a multiple of p by the assumption of the lemma. The only orbits with sizes that are not multiples of p are the fixed points and thus the claim follows. □

Now we can approach the second property in the theorem:

Lemma 1.9.6. *If $H \subset G$ is a finite p-subgroup and P is a subgroup of order p^K, then there exists an element $g \in G$ such that $gHg^{-1} \subset P$.*

Proof. Let X be the set of left cosets of P in G and let H act on X by left-multiplication. Applying Lemma 1.9.5 to H acting on X, we see that $|X| \equiv |X_0| \pmod p$. Note that p does not divide $(G : P)$, because P is of maximal order, so p does not divide $|X_0|$. Therefore in particular $X_0 \neq \emptyset$ and we pick $gP \in X_0$. It follows that for every $h \in H$ we have $hgP = gP$ and so $g^{-1}hgP \subset P$. In total we obtain $g^{-1}Hg \subset P$ as claimed. □

Note that this lemma implies that all subgroups of order p^K are conjugate, because we can apply the lemma twice to mutually embed conjugates of them into each other. Also $g^{-1}Hg \subset P$ implies $H \subset gPg^{-1}$ and thus every p-subgroup is contained in a subgroup of order p^K, i.e. in particular all p-Sylow subgroups are of order p^K. This proves the second statement.

In order to prove the last statement, note that Lagrange's corollary 1.7.4 implies $n_p = [G : N_G(P)]$ where P is any p-Sylow subgroup and $N_G(P)$ is the normalizer of P in G (because $\operatorname{Stab}_c(P) = N_G(P)$ is the stabilizer under conjugation action). Since $P \subset N_G(P) \subset G$ we have $n_p | m$ as stated in the theorem. Let $X = \operatorname{Syl}_p(G)$ the set of all p-Sylow subgroups of G and let P act on X by conjugation. Let $Q \in X_0$ with X_0 the fixed-point set. Note that this implies $gQg^{-1} = Q$ for all $g \in P$. Therefore $P \subset N_G(Q)$. By the last lemma P and Q are conjugate in G. Since Q is normal in $N_G(Q)$, we have $P = Q$. Therefore $X_0 = \{P\}$ and so by Lemma 1.9.5 the Theorem follows. □

Remark. It is not necessary to remember the whole proof for the exams. However the Theorem as well as some results of the lemmas in the proof can be helpful for proving things in the exam, the homework, or in research related to finite groups.

Example 1.9.7. Given a group of order 15 show that it is isomorphic to $\mathbb{Z}/(15)$.

The problem is of course to show that the group is abelian. We start with $15 = 3 \cdot 5$ and observe that $n_5 | 3$ and $n_5 \equiv 1 \pmod 5$ implies $n_5 = 1$. Therefore the only subgroup of order 5 is normal as it has no distinct conjugates. Similarly $n_3 | 5$ and $n_3 \equiv 1 \pmod 3$ imply $n_3 = 1$. Therefore also the only subgroup of order 3 is normal. Since the intersection of these two normal subgroups is $\{\mathrm{id}\}$, G must be a direct product of the groups of order 3 and 5. Therefore $G \cong \mathbb{Z}/(3) \times \mathbb{Z}/(5) \cong \mathbb{Z}/(15)$.

Corollary 1.9.8. *A finite group is a p-group iff every element has order a power of p.*

Proof. The Sylow theorem implies in particular that for $n = p_1^{n_1} \dots p_k^{n_k}$ with $p_i \in \mathbb{P}$ distinct primes and $n_i \in \mathbb{N}_+$, there are subgroups of order $p_i^{n_i}$. It is easy to see that such a subgroup must contain an element of order p_i^N for some integer $1 \leq N \leq n_i$. □

Remark 1.9.9. You could interchange the definition and corollary and say that a (possibly infinite) group G is a p-group if every element has order some (integer) power of p. Then it follows that a finite group is a p-group iff $(G : 1) = p^K$ (for some integer K). The only additional groups would be completely torsion, i.e. every element has finite order, but could still be non-abelian.

1.9.99 Exercises

Exercise 1.9.1. Given a finite group whose order is divisible by a prime p. Show that there is a subgroup of order p using Sylow's Theorem, but without using Lemma 1.9.4.

Exercise 1.9.2. Find the Sylow subgroups of

a. S_4 and

b. S_5.

Exercise 1.9.3. Find all groups of order

a. 33,

b. 35,

c. 45.

Exercise 1.9.4. Show that the following are not simple groups, i.e. that they have a non-trivial normal subgroup:

a. A group of order 18;

b. A group of order 30;

c. A group of order 56.

1.10 Small gods (分类的小群)

Given the structure theory we have been doing so far, you should try to work out a classification of all finite groups of order less than 60.

With Lagrange's theorem we know that every group of prime order is cyclic.

With Sylow's theorems we know that every group G of order pq where $p > q$ and $q \nmid (p-1)$ is abelian and hence cyclic, because $n_p \equiv 1 \pmod{p}$ and $n_p | q$ implies $n_p = 1$ as well as $n_q \equiv 1 \pmod{q}$ and $n_q | p$ implies $n_q = 1$ which are both normal and hence $G \cong \mathbb{Z}/(p) \times \mathbb{Z}/(q) \cong \mathbb{Z}/(pq)$.

Proposition 1.10.1. *A group of order $2p$ where p is a prime is either abelian or dihedral.*

Proof. Let G be the group in question. Remember that a group of order less than 6 is abelian. Let thus $p > 2$. We know that $n_p | 2$ and $n_p \equiv 1 \pmod{p}$ hence there is only one p-Sylow subgroup and it is of order p and normal. Conversely there are $n_2 | p$ and $n_2 \equiv 1 \pmod{2}$, 2-Sylow subgroups which are of order 2 each. Since $|G| = 2p = |\mathbb{Z}/(2)|\,|\mathbb{Z}/(p)|$ we know that $G = \langle a, b : a^2 = \mathrm{id} = b^p, \ldots \rangle$.

If $n_2 = 1$, we know that also the only subgroup of order 2 is normal and hence $G \cong \mathbb{Z}/(p) \times \mathbb{Z}/(2) \cong \mathbb{Z}/(2p)$. In general we have $aba = b^k$ for some $k \in \mathbb{N}_+$, because the p-Sylow subgroup is normal. From $a^2 = \mathrm{id}$ it also follows that $b = aaba^{-1}a^{-1} = ab^k a^{-1} = (aba^{-1})^k = b^{k^2}$, hence $p | (k^2 - 1)$. Since p is prime either $p | (k-1)$ or $p | (k+1)$.

If $p | (k-1)$, then $aba = b^k = b$ and thus $ab = ba$, i.e. G is abelian. If on the other hand $p | (k+1)$, then $aba = b^k = b^{-1}$ and thus $G \cong D_p$. □

p-groups

We already know that $(G : 1) = p$ implies that $G \cong C_p$, i.e. cyclic. For $k = 2$, i.e. $(G : 1) = p^2$, we have two obvious examples C_{p^2} and $(C_p)^{\times 2}$. Exercise 1.7.6 shows that G must be abelian as soon as we can show that the center is non-trivial. For $(G : 1) = p^2$ the center cannot be trivial, because of the class equation.

A bit more tricky is the situation of groups of order $8 = 2^3$. We have seen so far that D_4 and Q (the group of orthogonal unit quaternions) are non-abelian and not isomorphic.[9] We now claim that these are the only non-abelian possibilities.

Proposition 1.10.2. *Given a non-abelian group of order 8, then it is either isomorphic to D_4 or to Q.*

[9] e.g. by counting elements of order 2

Proof. Let G be the group in question. No element of G has order 8, because G is not cyclic. Also not every element can have order 1 or 2, because $g^2 = \mathrm{id}$ for all $g \in G$ would imply $gh = h^{-1}g^{-1} = hg$ for all $g, h \in G$, i.e. that G is abelian. Thus there is an element $a \in G$ of order 4 and $A := \langle a \rangle \triangleleft G$, because it has index 2. Therefore the group G is generated by a and any $b \in G \setminus A$, because $A \subsetneq \langle a, b \rangle \subseteq G$. Moreover $b^2 \in A$, because bA has order 2 in G/A. Also $b^2 \neq a^{\pm 1}$ for otherwise b had order 8. Hence $b^2 = \mathrm{id}$ or $b^2 = a^2$. Moreover $bab^{-1} \in A$ has order 4 like a, but $bab^{-1} \neq a$ (otherwise G were abelian), hence $bab^{-1} = a^3 = a^{-1}$.

If $b^2 = 1$, then the defining relations of D_4 hold in G, i.e. there is a homomorphism of D_4 onto G which is an isomorphism, because both groups have order 8. If on the other hand $b^2 = a^2$, then the defining relations of Q hold ….. This completes the proof. □

Remark 1.10.3. The principle for $n = 2^4$ or $n = 3^3$ is the same, i.e. first find all non-abelian examples (3 in the first case and 2? in the second case), and then try to prove that by exploiting the maximum order of an element in G.

Order $p^m q^n$

The next step are groups of order $p^m q^n$ where p and q are different primes. The abelian examples are obvious, but there are also plenty of non-abelian examples.

Let us consider groups of order 12. We have already seen D_6 and A_4 (which are not isomorphic). After some trying, we may also come up with $T := \langle a, b : a^6 = \mathrm{id}, b^2 = a^3, bab^{-1} = a^{-1} \rangle$. It seems that we cannot come up with further (non-isomorphic) examples. Indeed that can be shown to be correct starting with Sylow's theorem to obtain a normal subgroup, similarly to case of groups of order 8.

Proposition 1.10.4. *Every non-abelian group of order 12 is isomorphic to either D_6, A_4 or T.*

Idea of proof. Let G be the group in question. Since $12 = 3 \cdot 4$ we know that there is a subgroup $P \subset G$ of order 3. Then G acts by left-multiplication on the left-cosets G/P. Since $(G : P) = 4$ this is an action of G on a 4-element set, i.e. a homomorphism from G into S_4. Its kernel $K \triangleleft G$ is a normal subgroup of G. Moreover $K \subset P$, because $gxP = xP$ for all $x \in G$ implies $g \in P$. Therefore $K = 1$ or $K = P$.

If $K = 1$, then G is isomorphic to a subgroup $H \subset S_4$ of order 12. Let $\sigma \in S_4$ be any 3-cycle. Since $(S_4 : H) = 2$, two of 1, γ, γ^2 must be in the same left-coset of H, i.e. $\gamma \in H$ or $\gamma^2 \in H$, thus $\gamma = \gamma^4 \in H$, i.e. all 3-cycles are contained in H. Therefore $A_4 = H \cong G$.

order	type
1,2,3,5,7,11,13,15	cyclic,
4,9,	$\mathbb{Z}/(p^2)$ or $(\mathbb{Z}/(p))^{\times 2}$, all abelian,
6,10,14	cyclic or dihedral,
8	abelian, D_4 or Q,
12	abelian, D_6, A_4, or $T :=$ $\langle a, b : a^6 = \mathrm{id}, b^2 = a^3, bab^{-1} = a^{-1} \rangle$.

Table 1.1: Classification of groups of small order

If $P = K \triangleleft G$, then P is the only 3-Sylow subgroup of G. Thus G has only 2 elements of order 3. Thus every $c \in P$ has at most two conjugates and thus its centralizer (under the conjugation action) has order 6 or 12. Hence by Homework 1.9.1 there is an element $d \in \mathrm{cent}_G(c)$ of order 2. For $c \neq \mathrm{id}$, $a := cd$ has order 6. Define $A := \langle a \rangle$ and observe that $A \triangleleft G$, because it has index 2.

As in the proof of the last proposition G is generated by a and any $b \in G \setminus A$. Now $bab^{-1} \in A$ has order 6 like a. Moreover $bab^{-1} \neq a$ otherwise G were abelian, hence $bab^{-1} = a^5 = a^{-1}$. Also $b^2 \in A$, because bA has order 2 in G/A. $b^2 \neq a, a^5$, otherwise G were cyclic. Analogously $b^2 \neq a^2, a^4$, because b commutes with b^2 but $ba^2b^{-1} = a^{-2}$ yields $ba^2 = a^4b$. Hence $b^2 = \mathrm{id}$ which leads to D_6 or $b^2 = a^3$ which leads to T. This completes the proof. □

Remark 1.10.5. The cases $n = 18 = 2 \cdot 3^2$ or $n = 28 = 2^2 \cdot 7$ are correspondingly. The case $n = 24 = 2^3 \cdot 3$ is a bit harder (more examples, longer proof).

Orders containing 3 or more distinct primes

Remark 1.10.6. The smallest case is $n = 2 \cdot 3 \cdot 5 = 30$ and has only one non-abelian example.

The proofs get essentially more complicated from $n = 60 = 2^2 \cdot 3 \cdot 5$ on, because A_5 also has order 60 and no (non-trivial) normal subgroup (i.e. is simple), so the Sylow theorem(s) do not provide any normal subgroup.

Overview for order up to 15

In summary we have established Table 1.1.

1.10.99 Exercises

Exercise 1.10.1. To which group of order 12 in Table 1.1 is $C_2 \oplus D_3$ isomorphic?

Non-abelian groups in the following exercise should be specified by (easy) presentations.

Exercise 1.10.2. Find all groups of order

a. 51,

b. 21,

c. 39,

d. 55,

e. 57,

f. 93.

1.11 The general linear group (一般线性群)

Remember the notion of a vector space (向量空间) V over a field (域) F. The linear maps (线性图) from V to V can be composed and form thus a semigroup with neutral element (the identity $\mathbb{1}_V$). In order for an endomorphism to be invertible its determinant (行列式) may not vanish. The adjunct formula (伴随逆公式, $(\operatorname{cod} A)A = A \operatorname{cod} A = (\det A)\mathbb{1}$) then shows that indeed all such linear maps are invertible.

Definition 1.11.1. *Given a vector space V (over a field F), then the general linear group* $\mathrm{GL}(V)$ *is defined as the set of all linear maps $\phi\colon V \to V$ with non-vanishing determinant* $\det \phi \neq 0$.

Note that given a finite dimensional vector space by choosing a base $\{e_1, \ldots, e_n\}$ this is isomorphic to the standard vector space F^n. The linear maps are now encoded by $n \times n$-matrices with entries in F, denoted $\mathrm{Mat}_n(F)$. The set of all invertible matrices is an open subset, because the condition $\det A \neq 0$ is open. It is therefore reasonable to say that $\mathrm{GL}_n(F)$ as well as $\mathrm{GL}(V)$ has dimension n^2.

Example 1.11.2 (important subgroups)**.** 1. $\mathrm{SL}_n(F) := \ker \det = \{g \in \mathrm{GL}_n(F) : \det g = 1\}$ the volume preserving linear transformations (特殊线性群), obviously $\mathrm{GL}_n(F) = \mathrm{SL}_n(F) \cdot F^*$ the latter embedded e.g. via $F^* \hookrightarrow \mathrm{GL}_n(F) : \lambda \mapsto \mathrm{diag}(\lambda, 1, \ldots, 1)$. Note however that this does not commute, i.e. $\mathrm{GL}_n(F)$ is not a direct product;

2. $\mathrm{PSL}_n(F) := \mathrm{SL}_n(F)/\operatorname{cent}$ where the center of $\mathrm{SL}_n(F)$ is $\operatorname{cent} = \{\lambda \in F : \lambda^n = 1\}$. For $n \geq 3$ these groups are simple. It is also possible to define the analogue for GL as $\mathrm{PGL}_n(F) := \mathrm{GL}_n(F)/\operatorname{cent}$ where the center is correspondingly $\operatorname{cent} \mathrm{GL}_n(F) = \{\lambda \mathbb{1} : \lambda \in F^*\} \cong F^*$;

3. $\mathrm{O}(n) := \{g \in \mathrm{GL}_n(F) : g^T g = \mathbb{1}\}$ the orthogonal transformations (正交群), this is a maximal compact subgroup if F is real (i.e. $x^2 \geq 0$ for all $x \in F$), $\mathrm{SO}(n) := \mathrm{SL}_n(F) \cap \mathrm{O}(n)$ the orientation preserving (取向保存) orthogonal transformations;

4. $\mathbb{U}(n) := \{g \in \mathrm{GL}_n(\mathbb{C}) : gg^\dagger = \mathbb{1}\}$ the unitary transformations (酉群), $\mathbb{U}(n) \subset \mathrm{GL}_n(\mathbb{C})$ is correspondingly maximal compact, $\mathrm{SU}(n) := \mathrm{SL}_n(\mathbb{C}) \cap \mathbb{U}(n)$ the volume-preserving unitary transformations. Note that $\mathbb{U}(1) = \mathbb{S}^1$ the circle (圆). Also $\mathrm{SU}(2) = \mathbb{S}^3$ (3-dim'l sphere, 3D 超球面), which can be seen if you expand $gg^\dagger = \mathbb{1}$ for $\mathrm{GL}_2(\mathbb{C}) \ni g = \begin{pmatrix} a + b\mathsf{i} & c + c'\mathsf{i} \\ \dots & \end{pmatrix}$ as $0 = \dots$ (thus $c' = 0 = d'$) and $1 = a^2 + b^2 + c^2 + d^2$;

5. $\mathrm{SP}_n(F) := \{g \in \mathrm{GL}_{2n}(F) : g^T J g = J\}$ where $J = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}$, the symplectic transformations (辛群);

6. Note that $\mathbb{T}^n := \mathbb{U}(1)^n \hookrightarrow \mathrm{GL}_n(\mathbb{C})$ the (generalized) torus (环面) via diagonal elements. This is a maximal abelian subgroup. All maximal abelian subgroups are conjugate;

7. $B := \{\begin{pmatrix} * & * & \dots & * \\ 0 & * & *\dots & * \\ 0 & \dots 0 & \ddots & \vdots \\ 0 & \dots & 0 & * \end{pmatrix}\} \subset \mathrm{GL}_n(F)$ the upper triangular matrices, the Borel group, a maximally solvable subgroup. If the diagonal is restricted to 1s, then the group is even nilpotent;

8. $\mathrm{ISO}(n) \cong \mathrm{O}(n) \ltimes \mathbb{R}^n$, i.e. the elements are (R, v) where $R \in \mathrm{O}(n)$ is any isometry that fixes the origin and $v \in \mathbb{R}^n$ is a translation vector. The group operation is $(R, v)(R', v') = (RR', v + Rv')$ (with the neutral element $(\mathbb{1}, 0)$ and the inverse elements $(R, v)^{-1} = (R^{-1}, -R^{-1}v)$. This is called *semi-direct product (半直积*; it is not the direct product, but almost*)*. Note that $\mathrm{ISO}(n) \subset \mathrm{GL}_{n+1}(\mathbb{R})$.

Numerically interesting is the following decomposition:

$$\mathrm{GL}_n(\mathbb{C}) = \mathrm{P}_n^+(\mathbb{C})\mathbb{U}(n), \quad A = PU$$

where $\mathrm{P}_n^+(\mathbb{C})$ are the self-adjoint positive definite matrices. In the case $n = 1$ this reduces to $z = r\mathrm{e}^{\mathrm{i}\phi}$, i.e. the representation of a (nonzero) complex number in trigonometric form ($r > 0$ and $\mathrm{e}^{\mathrm{i}\phi} \in \mathbb{U}(1)$). The general case is called *polar decomposition (极分解)*. Analytically you can define $P := \sqrt{A^\dagger A}$ (where $A^\dagger$ is the adjoint matrix and thus $A^\dagger A$ self-adjoint and positive (semi)-definite) and $U := P^{-1}A$ (uniquely if A and thus P are invertible).

1.11.99 Exercises

Exercise 1.11.1. Determine the range of the determinant when restricted to the following subgroups

a. $\mathrm{O}(n)$,

b. $\mathbb{U}(n)$,

c. $\mathrm{SP}_n(F)$.

Exercise 1.11.2. Show that $\mathrm{O}(n)$ consists of at least 2 connected components. Conclude that the same is also true for $\mathrm{GL}_n(\mathbb{R})$.

1.12 Group representations (群表示论)

Definition 1.12.1. *Given a discrete group G, a representation of G on a vector space V is a group homomorphism $\rho\colon G \to \mathrm{GL}(V)$.*

Example 1.12.2.

0. Given any group G and any vector space V, we can map $\rho\colon G \to \mathrm{GL}(V) : g \mapsto \mathbb{1}$, i.e. all elements are mapped to the identity of V. This is called *a* trivial representation. More particularly $V = 0$ and thus $\mathrm{GL}(V) = \{\mathbb{1}\}$ gives *the trivial representation.*

1. Given a finite group G we define the vector space $F[G]$ with base $\{e_g : g \in G\}$ and the representation $\lambda\colon G \to \mathrm{GL}(F[G]) : g \mapsto (e_h \mapsto e_{gh})$, called the (left) regular representation (规则表示). As a more particular example consider $G = \mathbb{Z}/(3)$ and thus $V = F[\mathbb{Z}/(3)] \cong \mathbb{R}^3$,
$$\lambda(0) = \mathbb{1}, \quad \lambda(1) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda(2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

2. Given two representations $\pi_i\colon G \to \mathrm{GL}(V_i)$ of a group G, then we can construct a new representation on $V := V_1 \oplus V_2$, $\pi\colon G \to \mathrm{GL}(V_1 \oplus V_2) : g \mapsto \pi_1(g) \oplus \pi_2(g)$ where the matrices are written diagonally above each other.

A representation is now called *irreducible* (不可约表示) if it is nontrivial and cannot be written as the direct sum of two nontrivial representations. Conversely a representation is called *fully reducible* if it can be written as the direct product of irreducible representations.

3. Given two representations $\pi_i\colon G \to \mathrm{GL}(V)$ of a group G, then we can construct a new representation on $V := V_1 \otimes V_2$, $\pi\colon G \to \mathrm{GL}(V_1 \otimes V_2) : g \mapsto \pi_1(g) \otimes \pi_2(g)$. This way the representations (up to isomorphism) of a group form a *semi-ring (半环*, addition has no inverses either*)*. The other operations of vector spaces also generalize to representations, e.g. $\bigwedge^p \pi\colon G \to \mathrm{GL}(\bigwedge^p V) : g \mapsto \bigwedge^p \pi(g)$, and the like.

Given a real (complex) representation $\pi\colon G \to \mathrm{GL}(V)$ with a positive definite bilinear (sesquilinear) form $\langle .,. \rangle\colon V \otimes V \to F$ ($F = \mathbb{R}$ or $\mathbb{C}$), we can average the inner product as

$$\langle\!\langle v, w \rangle\!\rangle := \frac{1}{|G|} \sum_{g \in G} \langle \pi(g)v, \pi(g)w \rangle$$

(note that $\langle\!\langle .,. \rangle\!\rangle$ is still a positive definite inner product) and observe that now $\pi\colon G \to \mathrm{O}(V, \langle\!\langle .,. \rangle\!\rangle)$ (or $\mathbb{U}(V, \langle\!\langle .,. \rangle\!\rangle)$), i.e.

$$\langle\!\langle \pi(g)u, \pi(g)v \rangle\!\rangle = \frac{1}{|G|} \sum_{h \in G} \langle \pi(h)\pi(g)u, \pi(h)\pi(g)v \rangle = \frac{1}{|G|} \sum_{k \in G} \langle \pi(k)u, \pi(k)v \rangle = \langle\!\langle u, v \rangle\!\rangle.$$

In the case of fields with characteristic nonzero (i.e. $1/|G|$ might not be in F) representations with an invariant inner product are fully reducible because of the following fact.

Lemma 1.12.3. *Given an invariant subspace $U \subset V$ of a representation $\pi\colon G \to \mathrm{O}(V)$, i.e. $\pi(g)U \subset U$ for all $g \in G$, with an invariant inner product. Then V decomposes into two representations $V = U \oplus U^\perp$.*

Proof. The only thing that needs to be shown is that $U^\perp$ is also invariant under G. Note therefore the definition of $U^\perp = \{v \in V : \langle U, v \rangle = 0\}$ and let $v \in U^\perp$ be such a vector and $g \in G$. Then

$$\langle U, \pi(g)v \rangle = \langle \pi(g^{-1})U, v \rangle \subset \langle U, v \rangle = \{0\},$$

therefore also $\pi(g)v \in U^\perp$ and thus $\pi(g)U^\perp \subset U^\perp$. Since $\pi(g)$ is invertible and orthogonal also $\pi(g)|U^\perp$ is invertible and orthogonal. This completes the proof. □

Therefore fully reducible representations split uniquely into direct products of irreducible representations.

An interesting question is how to check whether a representation is irreducible. This can be decided with the following notion.

Definition 1.12.4. *Given two representations $\pi_i\colon G \to \mathrm{GL}(V_i)$. An equivariant map (等变映射) is a linear map $T\colon V_1 \to V_2$ such that $T \circ \pi_1(g) = \pi_2(g) \circ T$ for all $g \in G$.*

An intertwiner (交结映射) is an invertible equivariant map. (In particular the vector spaces must be of the same dimension.)

A self-intertwiner (自交结映射) of a representation is an intertwiner of the representation with itself.

Example 1.12.5. 0. Given any representation $\pi\colon G \to \mathrm{O}(V)$, then $\lambda \in \mathbb{F}^*$ gives the trivial self-intertwiners $\lambda\mathbb{1}$.

1. Given a decomposition into invariant subspaces $U_i \subset V$, $V = U_1 \oplus U_2$ of a representation $\pi\colon G \to \mathrm{O}(V)$, then $\lambda_i \in \mathbb{F}^*$ gives the (nontrivial) self-intertwiners $T\colon V \to V : u_1 \oplus u_2 \mapsto \lambda_1 u_1 \oplus \lambda_2 u_2$.

The statement is now named after I. Schur[10] and reads as follows.

Proposition 1.12.6 (Schur's lemma, 舒尔引理)**.** *Given a representation $\pi\colon G \to \mathrm{O}(V)$, then it is irreducible iff all self-intertwiners are multiples of the identity $\lambda\mathbb{1}$.*

Proof. In one direction, let $U \subset V$ be an invariant subspace. Therefore $V = U \oplus U^\perp$ and thus we can construct the non-trivial intertwiner T as in Example 1.12.5-1.

Let conversely $T\colon V \to V$ be a non-trivial self-intertwiner. Since the representation is orthogonal, its transpose $T^\dagger$ is also a self-intertwiner. By averaging we can assume that T is self-adjoint. Therefore we can diagonalize it with eigenvalues $\{\lambda_k : k = 1, \dots, n\}$ where there must be at least one eigenvalue. If there were only one eigenvalue, then T would be scalar, thus a trivial intertwiner. Therefore there must be at least 2 different eigenvalues. To each eigenvalue there is a T-invariant eigenspace $V = U_1 \oplus U_2 \oplus \dots$ containing at least one nonzero eigenvector each. Since T intertwines with the representation, the representation preserves the eigenspaces (otherwise the eigenvalue of a (generalized) eigenvector could change). Therefore we have an invariant subspace U_1 which is nontrivial. Together with the last lemma this gives a decomposition of V. This completes the proof. □

Example 1.12.7. Given the (left)-regular representation λ of a finite group G, then this has one invariant subspace $V_{\mathbb{1}} := \langle \sum_{g\in G} e_g \rangle$. This corresponds to the trivial representation $1\colon G \to \mathrm{GL}(F) : g \mapsto \mathbb{1}$.

Definition 1.12.8. *Given a finite dimensional representation $\pi\colon G \to \mathrm{GL}(V)$, we define its character (特征标) $\chi\colon G \to F^* : g \mapsto \operatorname{tr} \pi(g)$.*

[10] ∗1/1875 in Mogliev †1/1941

Proposition 1.12.9. *A character is a class function, i.e.* $\chi(ghg^{-1}) = \chi(h)$ *for all* $g, h \in G$. □

Remark 1.12.10. Note that for finite (compact) groups G the characters of complex representations are elements of $L_2(G)$ the Hilbert space over the group. The inner product is for finite groups

$$\langle \chi, \chi' \rangle := \frac{1}{|G|} \sum_{g \in G} \overline{\chi(g)} \chi'(g).$$

Example 1.12.11. Consider the following representations of S_3

π	χ : id	(12)	(123)	$\lvert\chi\rvert^2$	comment
λ	6	0	0	6	reg. repr.
1	1	1	1	1	$\langle \lambda, 1 \rangle = 1$
sgn	1	-1	1	1	$\langle \lambda, \mathrm{sgn} \rangle = 1$
π_3	3	1	0	2	fund. repr.
$\pi_2 := \pi_3 \ominus 1$	2	0	-1	1	$\langle \lambda, \pi_2 \rangle = 2$

And therefore $\lambda = 1 \oplus \mathrm{sgn} \oplus \pi_2 \oplus \bar{\pi}_2$.

Proposition 1.12.12. *Given a representation* $\pi \colon G \to \mathrm{GL}(V)$ *on a vector space over an algebraically closed field*[11] *and its character* χ*, then*

1. *The scalar product of two characters is a non-negative integer,*
2. π *is irreducible iff* $|\chi|^2 = 1$*,*
3. χ_1 *the character of an irreducible representation* $\pi_1 \colon G \to \mathrm{GL}(V_1)$*, then* π*s decomposition into irreducible representations contains exactly* $\langle \chi, \chi_1 \rangle$ *copies of* π_1*.*

Proof. It is sufficient to show part 2 together with $\langle \chi, \chi' \rangle = 0$ for characters of non-isomorphic irreducible representations, the rest then follows from the bilinearity of the scalar product. □

1.12.99 Exercises

Exercise 1.12.1. Given a finite subgroup of $G \subset \mathrm{GL}(V)$ a real (or complex) vector space, show that

a. every $g \in G$ has determinant $\det g \in \Omega_*$ the group of roots of unity 1,

[11] such as $\mathbb{C}$

b. give an example of an element in $g \in \mathrm{GL}_2(\mathbb{R})$ that has finite order, but not determinant 1,

c. G is isomorphic to a subgroup of $\mathrm{O}(V)$ (or $\mathbb{U}(V)$ respectively).

Exercise 1.12.2. Decompose the (left)-regular representation of $S_3 \times C_4$ over the complex numbers $\mathbb{C}$ into irreducible representations.

Exercise 1.12.3 (Representation induced by a finite group action)**.** a. Simplify the definition of a group action (Definition 1.7.1) on a finite space X in terms of a group homomorphism and a symmetric group.

b. Show that every action μ of G on a finite space X induces a representation on the finite-dimensional vector space $F[X] := \langle e_x : x \in X \rangle_F$.

c. What do orbits and fixed-points relate to?

1.13 Composition series (合成列) and Jordan–Hölder theorem (若尔当–赫尔德定理)

Analysis by normal series is another powerful tool to study finite groups.

Remember that a group G is called *simple* iff it has no proper normal subgroups (i.e. normal subgroups other than 1 and G).

Definition 1.13.1. *Given a group G, then a* normal series (正规列) *for G is a series $G_0 = \{\mathrm{id}\} \triangleleft G_1 \triangleleft \ldots G_n = G$.*

A composition series (合成列) *is a normal series where all factors are simple.*

Example 1.13.2. 1. Consider the group $G = S_4$ with the normal series $\{\mathrm{id}\} \triangleleft \langle (12)(34) \rangle \triangleleft V_4 \triangleleft A_4 \triangleleft S_4$. This is a normal series of length 4. Note that the factors are all cyclic of prime order, thus it is a composition series.

2. Consider on the other hand the normal series $\{\mathrm{id}\} \triangleleft A_5 \triangleleft S_5$. It is rather short and A_5 is not abelian. Nevertheless it is simple and the series thus a composition series.

3. Consider $\mathbb{Z}/(6)$. It has two composition series, namely $0 \triangleleft \mathbb{Z}/(2) \triangleleft \mathbb{Z}/(6)$ and $0 \triangleleft \mathbb{Z}/(3) \triangleleft \mathbb{Z}/(6)$. But the factors G_k/G_{k-1} are $(\mathbb{Z}/(2), \mathbb{Z}/(3))$ and $(\mathbb{Z}/(3), \mathbb{Z}/(2))$, respectively, thus isomorphic up to order. This happens for all composition series of a fixed group.

Theorem 1.13.3 (Jordan[12]–Hölder[13]). *Given two composition series* $\{\mathrm{id}\} \lhd G_1 \lhd \ldots \lhd G_m = G$ *and* $\{\mathrm{id}\} \lhd H_1 \lhd \ldots \lhd H_n = G$ *of the same group* G*, then there is a bijection between the factors such that corresponding factors are isomorphic. (In particular* $m = n$*.)*

Proof (Schreier[14]). The main idea of this short version of the proof came from O. Schreier. We show that given any two normal series $\{\mathrm{id}\} \lhd A_1 \lhd A_2 \lhd \ldots \lhd A_m = G$ and $\{\mathrm{id}\} \lhd B_1 \lhd B_2 \lhd \ldots \lhd B_n = G$, then there are refinements $\{\mathrm{id}\} \lhd C_1 \lhd C_2 \lhd \ldots \lhd C_{mn} = G$ of $A_\bullet$ and $\{\mathrm{id}\} \lhd D_1 \lhd D_2 \lhd \ldots \lhd D_{mn} = G$ of $B_\bullet$ together with a permutation σ of $\{1, 2, \ldots, mn\}$ such that $C_k/C_{k-1} \cong D_{\sigma k}/D_{\sigma(k-1)}$.

The idea is to define

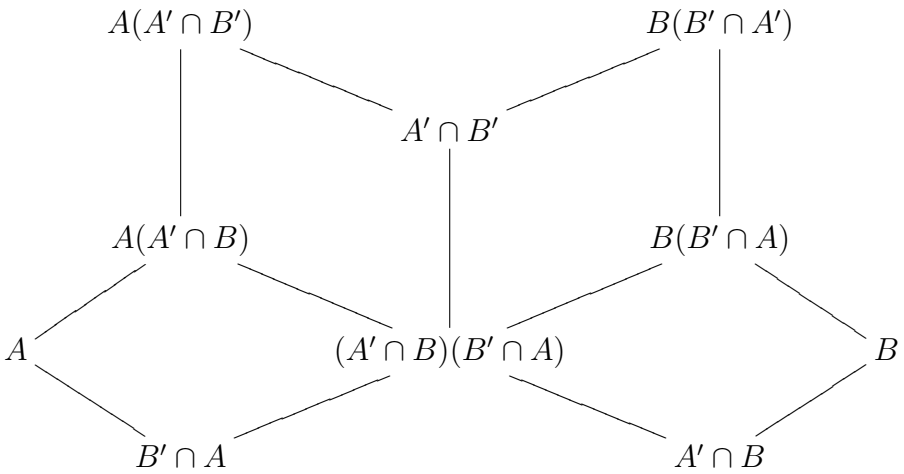
$$C_{ni+j} := A_i(A_{i+1} \cap B_j), \quad D_{mj+i} := B_j(B_{j+1} \cap A_i),$$

where $1 \le i \le m$ and $1 \le j \le n$. Obviously $A_i = C_{ni} \subset C_{ni+1} \subset \cdots \subset C_{n(i+1)} = A_{i+1}$ and the analogue for $B_j = D_{mj}$. The main part is to show that $C_\bullet$ and $D_\bullet$ are normal series. This derives from the following lemma.

Lemma 1.13.4 (butterfly $\sim$, Zassenhaus,[15] 蝴蝶引理). *If* $A \lhd A' \subset G$ *and* $B \lhd B' \subset G$*, then* $A(A' \cap B)$*,* $A(A' \cap B')$*,* $B(B' \cap A)$ *and* $B(B' \cap A')$ *are subgroups of* G*,* $A(A' \cap B) \lhd A(A' \cap B')$ *and* $B(B' \cap A) \lhd B(B' \cap A')$*, and*

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

The subgroup inclusion pattern is



[12] M.E.C. Jordan *1/1838 in Lyon/France, †1/1922
[13] O.L. Hölder *12/1859 in Stuttgart/ Germany, †8/1937
[14] O. Schreier *3/1901 in Vienna, Austria, †6/1929
[15] H.J. Zassenhaus *5/1912 in Germany, †11/1991

Proof. $A(A' \cap B)$ and $A(A' \cap B')$ are both subgroups of A', because $A \lhd A'$. Also $A' \cap B \lhd A' \cap B'$, because $B \lhd B'$.

Let now $x = ab' \in A(A' \cap B')$ and $y = a_2 b \in A(A' \cap B)$ with $a, a_2 \in A$, $b' \in A' \cap B'$ and $b \in A' \cap B$. Then $xa_2x^{-1} \in A$, because $x \in A'$ and $A \lhd A'$. Analogously $b'bb'^{-1} \in A' \cap B$, because $b' \in A' \cap B'$ and $A' \cap B \lhd A' \cap B'$. Moreover $xbx^{-1} = ab'bb'^{-1}a^{-1} \in A(A' \cap B)A = A(A' \cap B)$, because $A \lhd A'$. Finally $xyx^{-1} = xa_2x^{-1}xbx^{-1} \in A(A' \cap B)$. This shows $A(A' \cap B) \lhd A(A' \cap B')$.

Let furthermore $S := A' \cap B'$, $T := A(A' \cap B)$, and $U := A(A' \cap B')$. Then $S \subset U$ and $T \lhd U$. Therefore

$$ST = TS = A(A' \cap B)(A' \cap B') = A(A' \cap B') = U.$$

Next we need to determine $S \cap T$. Starting with $A \cap B' \subset S$, $A \cap B' \subset A \subset T$, $A' \cap B \subset S$, and $A' \cap B \subset T$, we note that $(A \cap B')(A' \cap B) \subset S \cap T$. Conversely for every $t \in S \cap T$, we observe $t \in A' \cap B'$ and so $t = ab$ for some $a \in A$ and $b \in A' \cap B$. But then $b \in B'$, $a = tb^{-1} \in B'$, and $t = ab \in (A \cap B')(A' \cap B)$. Therefore $S \cap T = (A \cap B')(A' \cap B)$ and in particular

$$A(A' \cap B')/A(A' \cap B) = ST/T \cong S/(S \cap T) = (A' \cap B')/(A \cap B')(A' \cap B).$$

Exchanging the As and Bs in the first argument yields that $B(B' \cap A)$ and $B(B' \cap A')$ are subgroups of G each, $B(B' \cap A) \lhd B(B' \cap A')$, and

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

Therefore $A(A' \cap B')/A(A' \cap B) \cong B(B' \cap A')/B(B' \cap A)$. This completes the proof of the lemma. □

With the choice $A = A_i$, $A' = A_{i+1}$, $B = B_j$ and $B' = B_{j+1}$ we see that $C_{ni+j} = A(A' \cap B) \lhd A(A' \cap B') = C_{ni+j+1}$, analogously for D_{mj+i}, and

$$C_{ni+j+1}/C_{ni+j} = A(A' \cap B')/A(A' \cap B) \cong B(B' \cap A')/B(B' \cap A) = D_{mj+i+1}/D_{mj+i}.$$

Moreover, $C_{ni} = A_i$ and $D_{mj} = B_j$. Therefore $C_\bullet$ is a refined normal series of $A_\bullet$, $D_\bullet$ a refined normal series of $B_\bullet$, and $C_\bullet$ and $D_\bullet$ have the same factors up to the rearrangement $\sigma\colon \{1, \dots, mn\} \to \{1, \dots, mn\} : ni + j \mapsto mj + i$ for all $1 \le i \le m$ and $1 \le j \le n$, because $C_k/C_{k-1} \cong D_{\sigma k}/D_{\sigma(k-1)}$ by the definition of $C_\bullet$ and $D_\bullet$. It is easy to see that $\sigma \in S_{mn}$ is indeed a permutation. Since $A_\bullet$ and $B_\bullet$ are already composition series in the beginning, the $C_\bullet$ and $D_\bullet$ only contain additional trivial steps, i.e. $C_k/C_{k+1} \cong \{1\}$. By omitting these factors we see that σ reduces to a bijection of the original simple factors of $A_\bullet$ and $B_\bullet$. This completes the proof of the Jordan–Hölder theorem. □

Note that the proof also works for normal series (which are not necessarily composition series) and gives then joint refinements.

Remember that a group G is said to be of *finite length* iff every chain of subgroups $\{1\} \subset A_1 \subset A_2 \subset \cdots \subset G$ (ascending) and $G \supset A_1 \supset A_2 \supset \cdots \supset \{1\}$ (descending) finally becomes constant. It is clear the every finite group is of finite length, but also some infinite groups such as ... are of finite length, because it has only finitely many subgroups ($\{1\}$, ...).

Proposition 1.13.5. *Given a group of finite length, then it has a composition series.*

Proof. Let G denote the group. We start with the trivial series $\{1\} \triangleleft G$ and refine it as long as any of the factors A_k/A_{k+1} is not yet simple. Since the group G has finite length, this process must stop after finitely many steps. Since each factor is simple, there is no further refinement and the obtained series thus a composition series. □

We therefore see that finite groups are built of simple groups. A first step in a classification/ complete understanding of all finite groups is thus to understand the simple groups. This has indeed been achieved[16] and leads to 18 infinite series of simple groups together with 26 exceptions of simple groups. Unfortunately the proof is some 2100 pages long.

Example 1.13.6. 0. The easiest simple groups are the cyclic groups of prime order $\mathbb{Z}/(p)$ for every prime $p \in \mathbb{P}$. Namely an abelian group is simple iff it is cyclic of prime order. Conversely every finite abelian group is the direct sum of cyclic groups of order a power p^k of a prime p (see Theorem 1.5.7). And the latter have composition series of length k all factors being C_p.

1. Another family we already know are the alternating groups A_n for $n \geq 5$.

2. Yet another series are the $\mathrm{PSL}_n(F)$ where F is a finite field, $n \geq 3$, $\mathrm{SL}_n(F)$ are the linear isomorphisms of F^n of determinant 1, and $\mathrm{PSL}_n(F) := \mathrm{SL}_n(F)/\operatorname{cent}$. The center consists of $\{\lambda \mathbb{1} : \lambda \in F^*, \lambda^n = 1\}$.

1.13.1 Group extensions (群扩张)

The harder question is how to reconstruct an arbitrary group from its composition factors. The elementary step is the second part of the following definition.

Definition 1.13.7. *A sequence of group homomorphisms* $\cdots \to G_{n-1} \xrightarrow{\phi_{n-1}} G_n \xrightarrow{\phi_n} G_{n+1} \to \ldots$ *is called* exact (正合序列) *if at every group* $\ker \phi_n = \operatorname{im} \phi_{n-1}$.

An exact sequence $\{\mathrm{id}\} \to N \to G \to Q \to \{\mathrm{id}\}$ *is called* short exact sequence (短正合序列) *or* extension (群扩张) *of the group* Q *by* N.

[16] Gorenstein et al, ∼ 2100pp, **(1994)**.

Example 1.13.8. 0a The beginning $1 \to S \to G$ means that S is embedded into G, i.e. a subgroup (up to isomorphism),

0b The ending $G \to Q \to 1$ means that G maps surjectively onto Q.

1. Both together $1 \to G \to Q \to 1$ mean that $G \cong Q$, i.e. the two groups are isomorphic.

2. Consider the sequence $1 \to C_2 \to G \to C_2 \to 1$ and let us ask how many inequivalent groups G exist here. Since $N = C_2 = Q$ are of order 2 each, $|G| = 4$ and thus G is abelian. But for abelian groups the classification is easy. The candidates are C_4 and $C_2 \times C_2$ and by inspection both are possible group extensions.

1.13.99 Exercises

Exercise 1.13.1. a. Show that D_4 has a normal series where one of the components is not a normal subgroup of D_4.

b. Given normal series for $N \triangleleft G$ and G/N. Show that these can be pieced together to give a normal series of the group G.

c. Let $\{\mathrm{id}\} \triangleleft G_1 \triangleleft G_2 \triangleleft \ldots \triangleleft G_n = G$ be a normal series. Explain how normal series of each factor G_k/G_{k-1} give a refinement of this normal series. What is needed to obtain a composition series?

d. Given composition series of $N \triangleleft G$ and G/N. How can you obtain a composition series for the group G?

Exercise 1.13.2. If G has a composition series, then every normal subgroup $N \triangleleft G$ and every quotient G/N (by a normal subgroup) has a composition series.
Hint: Show how N appears in a composition series.

Exercise 1.13.3. Find all composition series of

a. A_4,

b. D_4,

c. D_5.

Exercise 1.13.4. Show that all abelian groups of order n have the same simple factors.

Exercise 1.13.5. Show that the simple factors of D_n are all abelian. (This means that D_n is solvable.)

Exercise 1.13.6. Let G be a group of order n and m the length of each of its composition series.

a. Show that a group of order $n = p^m$ where $p \in \mathbb{P}$ is a prime and $m \in \mathbb{N}_+$ a positive integer has a composition series of length m.

b. Show that $m \leq \log_2 n$.

c. Show that equality $m = \log_2 n$ is possible for arbitrary high values of n.

Exercise 1.13.7. Find all group extensions in the following cases

a. of $Q = C_2$ by $N = C_3$,

b. of $Q = C_3$ by $N = C_2$,

1.14 Solvable groups (可解群)

The simplest examples of abelian and non-abelian groups are captured by the following definition.

Definition 1.14.1. *Given a group G it is called* solvable *if it has a finite normal series with only abelian factors.*

Example 1.14.2.
1. Remember the composition series of S_4, $\{\mathrm{id}\} \triangleleft \langle (12)(34) \rangle \triangleleft V_4 \triangleleft A_4 \triangleleft S_4$. This implies that S_4 is solvable.

2. On the other hand the composition series of S_5, $\{\mathrm{id}\} \triangleleft A_5 \triangleleft S_5$ proves that S_5 is not solvable (because A_5 is simple but not abelian).

3. Note that $\mathbb{Z}$ has no finite composition series. Nevertheless $\mathbb{Z}$ is abelian itself and thus solvable.

Remark 1.14.3. The name is historically and also coincides with a property of the Galois group of a field extension/ polynomial that is *solvable by radicals (可解用根式)*. Solvable groups are sometimes also called *metabelian (元可换群)*, suggesting that they are close to abelian groups.

Definition 1.14.4. *Given a group G, then the* commutator (整流子) *of two elements $a, b \in G$ is $[a, b] := aba^{-1}b^{-1}$. The* commutator subgroup (换位子群) *of G is $G' := [G, G] := \langle [a, b] : a, b \in G \rangle$.*

Example 1.14.5. 0. Given any group G, then $G' = \{\mathrm{id}\}$ iff G is abelian.

1. On the other hand given a simple non-abelian group G, then $G' = G$, because $G' \neq \{\mathrm{id}\}$ and $G' \triangleleft G$ is a normal subgroup (see also the next proposition).

These two extreme cases suggest that the commutator may somehow measure solvability.

Proposition 1.14.6. *Given a group G, then G' is a normal subgroup. Every group homomorphism $\phi\colon G \to A$ into an abelian group factors through $\pi\colon G \to G/G'$.*

Proof. G' is a subgroup, because we permit finite products of commutators and $[a,b]^{-1} = (aba^{-1}b^{-1})^{-1} = bab^{-1}a^{-1} = [b,a]$ for all $a, b \in G$. Moreover G' is invariant under conjugation, because

$$x[a,b]x^{-1} = xaba^{-1}b^{-1}x^{-1} = (xax^{-1})(xbx^{-1})(xax^{-1})^{-1}(xbx^{-1})^{-1}$$

for all $x \in G$. Therefore $G' \triangleleft G$. Let now $\phi\colon G \to A$ be any homomorphism into any abelian group A. Then obviously $0 = \phi(aba^{-1}b^{-1})$ and thus $aba^{-1}b^{-1} \in \ker\phi$. Therefore $G' \subset \ker\phi$ and thus ϕ factors through $\pi\colon G \to G/G'$. □

The *commutator (换位子序列)* or *derived series (导出列)* of a group G is

$$\ldots \triangleleft \mathrm{D}^{n+1}G := [\mathrm{D}^nG, \mathrm{D}^nG] \triangleleft \mathrm{D}^nG \triangleleft \mathrm{D}^{n-1}G \triangleleft \ldots \triangleleft \mathrm{D}G := [G,G] \triangleleft \mathrm{D}^0G := G.$$

If some derivation $\mathrm{D}^nG = \{\mathrm{id}\}$ is trivial, then this is indeed a normal series. In the last proposition we have shown that the factors D^{n-1}/D^n are all abelian, and thus G is solvable if the derived series terminates in 1. Conversely, if G is not solvable, then it has some non-abelian factor in every normal series, including every truncation of the derived series (filled with $\{\mathrm{id}\}$ on the left end). Therefore the full derived series cannot end in $\{\mathrm{id}\}$. We have thus arrived at the following property:

Proposition 1.14.7. *A group is solvable iff its derived series ends in $\{\mathrm{id}\}$.* □

Proposition 1.14.8. *The class of solvable groups is closed under the following operations:*

1. *Subgroup, i.e. if G is solvable, then so is every subgroup $H \subset G$.*
2. *Quotient, i.e. if G is solvable and $N \triangleleft G$ a normal subgroup, then also G/N is solvable.*
3. *Composition, i.e. if $N \triangleleft G$ is solvable and G/N is solvable, then so is G.*

Proof. These are elementary properties of normal series. Given a normal series $\{\mathrm{id}\} \triangleleft G_1 \triangleleft G_2 \triangleleft \ldots \triangleleft G_n = G$, then any subgroup H inherits a normal subseries

$H_k := H \cap G_k$ which remains normal (in H_{k+1}), because G_k is normal. By isomorphism theorem the factors H_k/H_{k-1} are isomorphic to subgroups of G_k/G_{k-1} and remain thus abelian.

Given a normal subgroup $N \triangleleft G$, then G/N inherits a normal factor series $N_k := (G_k N)/N$ which again by isomorphism theorem is a normal series and has factors N_k/N_{k-1} isomorphic to factors of the G_k/G_{k-1}.

Finally we can join the normal series of N and G/N to obtain a normal series of G that has the same factors, because $(G/N)_k/(G/N)_{k-1} \cong G_{n+k}/G_{k+k-1}$ for $G_{n+k} = \pi^{-1}(G/N)_k$ with $\pi\colon G \to G/N$ (by isomorphism theorem) and $N_n = N$. □

Remember that a p-group is a group G where every element $g \in G$ has order some integer power of the given $p \in \mathbb{P}$. In particular every finite p-group has order a power of p.

Proposition 1.14.9. *Every finite p-group is solvable.*

Proof. Let G denote the group in question and $|G| = p^n$ for some $n \in \mathbb{N}$. If $n \leq 1$ then G is cyclic and thus solvable. For $n \geq 2$ we note that there is a subgroup $N \subset G$ of order p^{n-1} which is normal in G (because it is either unique or G is a direct product of several p groups each of which is normal in G). But then an induction shows that N as well as $G/N \cong \mathbb{Z}/(p)$ are both solvable and by the last proposition so is G. □

Further examples of (finite) solvable groups arise from the following property:

Proposition 1.14.10. *Given a group G of order $p^m q$ where $p, q \in \mathbb{P}$ are primes and $m \in \mathbb{N}$ is a non negative integer, then G is solvable.*

Proof. We may assume that $p \neq q$. Let S be a p-Sylow subgroup of G. If $S \triangleleft G$, then we are done, because the last proposition shows that S is solvable and the last factor G/S is cyclic.

Let thus S not be normal in G. This means in particular that $S \subset N_G(S) \subsetneq G$, i.e. the normalizer of S (in G) is strictly smaller than G. Since $(G : S) = q$ a prime, this implies $N_G(S) = S$ and the number of p-Sylow subgroups in G is $n_p = (G : N_G(S)) = q$. Since the p-Sylow subgroups intersect trivially, i.e. $S \cap T = \{\mathrm{id}\}$ for any two of them, there are at least $q(p^m - 1)$ elements of order some positive power of p. That leaves only q elements in G of order some power of q and therefore there is only one q-Sylow subgroup $Q \subset G$ which must therefore be normal. But then $Q \cong \mathbb{Z}/(q)$ as well as G/Q of order p^m are solvable and thus G is solvable. □

Remark 1.14.11. An even stronger result is Burnside's[17] $p^m q^n$-Theorem which states that every group of order $p^m q^n$ where $p, q \in \mathbb{P}$ and $m, n \in \mathbb{N}$ integers is solvable. To prove that you need, e.g. some strong ring theory which will however be beyond the reach of this course.

Theorem 1.14.12 (Feit–Thompson)**.** *Given a finite group of odd order, then it is solvable.*

The proof is a part of the 2100 pages book series.[16]

1.14.99 Exercises

Exercise 1.14.1. What can you say if $DG = G$ for a group G?
Hint: Consider $D(A_5 \times A_5)$ and note that this direct product is not simple.

1.15 Nilpotent groups (冪零群)

Definition 1.15.1. *A group is* nilpotent *iff its* lower central series (降/下中心列) *ends in* $\{\mathrm{id}\}$*. The lower central series is defined as*

$$G_0 := G, \quad G_{n+1} := [G, G_n]$$

where $[a, b] := aba^{-1}b^{-1}$ *for* $a, b \in G$ *is the commutator of two group elements.*

Example 1.15.2. Consider the Borel group $B \subset \mathrm{GL}_n(F)$ of upper triangular matrices. It is nilpotent, because G_0 has the diagonal with 1s, G_1 has the first sub-diagonal all 0, G_2 has the first two sub-diagonals all 0, …, the last group $G_{n-1} = \{\mathbb{1}\}$.

Proposition 1.15.3. *A group* G *is nilpotent iff its* upper central series (升/上中心序列) *ends in* G*. The upper central series is defined as*

$$\gamma^0 := \{\mathrm{id}\}, \quad \gamma^{n+1} := \mathrm{cent}_G(\gamma^n).$$

1.15.99 Exercises

Exercise 1.15.1. Prove the proposition about the upper central series, i.e. a group is nilpotent iff its upper central series $Z_{n+1} := \{z \in G : \forall g \in G : [g, z] \in Z_n\}$ with $Z_0 := \{\mathrm{id}\}$ ends in G.
Hint: Suppose G is nilpotent of length n (i.e. $G_n = 1$), show that $Z_k \supset G_{n-k}$ and thus $Z_n = G_0 = G$. In the other direction show that $Z_n = G$ implies $G_k \subset Z_{n-k}$.

[17] W. Burnside *7/1852 in London, †8/1927

1.16 Semidirect products (半直积)

Definition 1.16.1. *Given a short exact sequence of groups* $\{\mathrm{id}\} \to N \hookrightarrow G \twoheadrightarrow Q \to \{\mathrm{id}\}$ *where in addition we have an (injective) group homomorphism* $s\colon Q \to G$ *that is a right-inverse of the projection* $\pi\colon G \to Q$ *(*$\pi \circ s = \mathrm{Id}_Q$*). This is called a* semidirect product.

Proposition 1.16.2. *The above situation is called a* right-splitting (右分裂) *of the short exact sequence. This is equivalent to the following so called* left-splitting (左分裂)*: There is a fiberwise surjective map* $p\colon G \twoheadrightarrow N$*, i.e.* $p\colon \pi^{-1}(q) \xrightarrow{\sim} N$ *for all* $q \in Q$*, that is a left-inverse of the embedding* $N \to G$ *(*$p \circ e = \mathrm{Id}_N$ *where* $e\colon N \hookrightarrow G$*).*

Note that p may in general fail to be a group homomorphism, namely p is a group homomorphism iff $Q \triangleleft G$, i.e. $G = N \times Q$ a direct product.

Proof. Let $e\colon N \hookrightarrow G$ and $\pi\colon G \twoheadrightarrow Q$ denote the maps in the short exact sequence. Given a right-splitting $s\colon Q \to G$, then for every $g \in G$, we can construct the element $\hat{g} := g \cdot s(\pi(g))^{-1}$. Since s and π are group homomorphisms, we observe $\pi(\hat{g}) = \pi(g) \cdot \pi(g)^{-1} = \mathrm{id} \in Q$ and thus there is an element $n \in N$ such that $e(n) = \hat{g}$. Again π and s group homomorphisms implies that $\ker \pi = \{\hat{g} : g \in G\} \triangleleft G$ is a subgroup which is isomorphic to N via e and maps $n \mapsto \hat{g}$. Therefore $p = e^{-1} \circ m(\mathrm{Id}, s \circ \pi)\colon G \twoheadrightarrow N : g \mapsto n$ is a fiberwise surjective map.

Conversely, given $p\colon G \twoheadrightarrow N$ fiberwise surjective and a left-inverse of e, then we can for every $h \in Q$ choose the unique $g \in G$ with $\pi(g) = h$ and $p(g) = \mathrm{id} \in N$. For fixed π and h the coset class $ge(N)$ is unique and $e \circ p = \mathrm{Id}_N$. **MG:** … Therefore $s\colon Q \to G : h \mapsto g$ is a group homomorphism. This completes the proof. □

Example 1.16.3. Consider the group $\mathrm{GL}_n(F)$ together with the fundamental representation on the vector space F^n. Considering the latter as abelian group, we want to find a group $G \approx \mathrm{GL}_n(F) \times F^n$ fitting into $\{0\} \to F^n \to G \to \mathrm{GL}_n(F) \to \{\mathbb{1}\}$. We choose the group law $(A, v)(B, w) := (AB, v + Aw)$. One can see easily that this is associative, has the neutral element $(\mathbb{1}, 0)$ and the inverse elements $(A, v)^{-1} = (A^{-1}, -A^{-1}v)$. Moreover F^n is invariant under the action of $Q = \mathrm{GL}_n(F)$ on G. Therefore $F^n \triangleleft G$ as well as $G/F^n \cong \mathrm{GL}_n(F)$. Moreover $s\colon \mathrm{GL}_n(F) \to G : A \mapsto (A, 0)$ is a group homomorphism and a right-inverse of π.

Note that the map $p\colon G \to F^n : (A, v) \mapsto v$ is ("intuitive" and) fiberwise bijective, but not a group homomorphism, i.e. for $Aw \neq w$ we have $p((A, v)(B, w)) = p(AB, v + Aw) = v + Aw \neq v + w = p(A, v)p(B, w)$.

Remark 1.16.4. Remember the short exact sequence $0 \to C_2 \to C_4 \to C_2 \to 0$. Even though $C_4 \twoheadrightarrow C_2$, there is no group homomorphism inverting this projection, because C_4 is not the direct product $C_2 \times C_2$.

Q: What is the general structure of semi-direct products?

Corollary 1.16.5. *A group G is a semi-direct product of the subgroups N and Q iff $N \triangleleft G$ is a normal subgroup, $N \cap Q = 1$, and $NQ = G$. In particular this implies that there is a unique action $\rho\colon Q \to \mathrm{Aut}(N), \rho(q)n = qnq^{-1} \in N \subset G$ for all $q \in Q$ and $n \in N$.*

Proof. This follows immediately from the short exact sequence ($N \triangleleft G$) and the right-splitting ($Q \subset G$ and $NQ = G$). □

1.16.99 Exercises

Exercise 1.16.1. Which of the extensions $1 \to C_3 \to G \to C_2 \to 1$ are semi-direct products?

Exercise 1.16.2. Show that the D_n are semi-direct products. What is the group action?

Chapter 2

Rings and algebras (环理论与代数, 3 weeks)

Three Rings for the Elven-kings in under the sky,
Seven for the Dwarf-lords in the halls of stone,
Nine for Mortal Men doomed to die.
And one for the dark lord on his dark throne
In the land of Mordor where the Shadows lie.

One ring to rule them all, One ring to find them,
One ring to bring them all and in the darkness bind them.
In the land of Mordor where the Shadows lie.[1]

2.1 Definition and examples

Definition 2.1.1. *A* (commutative) ring (环; with unit 1) *is a set R together with two operations $+\colon R \times R \to R$ and $\cdot\colon R \times R \to R$ that make $(R, +)$ an abelian group with neutral element* 0*, $(R, \cdot)$ an (abelian) semi-group with neutral element* 1*, compatible in the sense*

$$ba = ab \tag{2.1}$$

$$a(b + c) = ab + ac \tag{2.2}$$

Example 2.1.2.

1. The ring of integers is the set $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$ together with the addition $+$ and multiplication $\cdot$.
2. The cyclic groups $\mathbb{Z}/(n)$ where $n > 0$ is a positive integer, inherit a multiplication, because $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$ imply $ac \equiv bd \pmod{n}$.

[1] J.R.R. Tolkien "The Lord of the Rings"

Note that some of these rings have zero-divisors, e.g. $n = 6$, $2 \cdot 3 \equiv 0 \pmod 6$.

3. Let R be any ring and consider the polynomials in one variable denoted $R[x]$. These are the finite sequences under the operation of component-wise addition and polynomial multiplication, i.e.

$$(a_0 + a_1 x + \cdots + a_m x^m)(b_0 + b_1 x + \cdots + b_n x^n) = \sum_{k=0}^{m+n} \sum_{l=\max(0,k-n)}^{\min(m,k)} a_l b_{k-l} x^k$$

Since $R \subset R[x]$ is compatible with the original structure (subring), the neutral element is $0 \in R$, the unit is $1 \in R$ and we denote $\deg(a_0 + \cdots + a_n x^n) := n$ if $a_n \neq 0$ the degree of the polynomial. The degree of the 0-polynomial is $-\infty$.

4. Further rings are the rational $\mathbb{Q}$, real $\mathbb{R}$, and complex numbers $\mathbb{C}$, and these are not only rings, but also *fields (域).*

5. Given two rings R and S, we can consider their direct sum (product) $R \oplus S$ with component-wise operation, i.e. $(a_1, b_1)(a_2, b_2) = (a_1 a_2, b_1 b_2)$ and correspondingly for addition. The neutral elements are obviously $0 = (0, 0)$ and $1 = (1, 1)$, respectively. Also the inverses are $-1 = (-1, -1)$ or $-(a, b) = (-a, -b)$, in general. We can also do that with more than 2 rings. Note that the corresponding rings have plenty of zero-divisors, even if the factors are integral (i.e. have no zero-divisors).

A very similar notion is that of an algebra. This is defined as follows:

Definition 2.1.3. *An* (associative) algebra (代数) *over a field F is a vector space A/F together with an F-bilinear operation $\cdot\colon A \otimes A \to A$ (that is associative).*

Rings are called commutative unital algebras over $\mathbb{Z}$.

Example 2.1.4. 0. Note that the rings, e.g. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, are (unital commutative) $\mathbb{Z}$-algebras.

1. Given a vector space V over a field F, then its endomorphisms $\mathrm{End}(V) := \{(\phi\colon V \to V) : \text{linear}\}$ are not only a vector space over F, but indeed a unital algebra under matrix multiplication. A more particular example are the $n \times n$-matrices with entries in F denoted as $\mathrm{Mat}_n(F)$. Note that this algebra is not commutative. It is therefore not a commutative ring.

2. The direct sum (product) also works for algebras.

   In this way the endomorphisms of a vector space have two multiplicative structures, one as endomorphisms (composition) and another one as a vector space with a particular identification to F^{n^2} (thus being a direct sum of copies of the base field).

3. Remember the quaternions $\mathbb{H} = \mathbb{R}(i, j, k)$. This is not a commutative ring, because $wz \neq zw$ for arbitrary $w, z \in \mathbb{H}$. Nevertheless it is a vector space over $\mathbb{R}$ and the multiplication is $\mathbb{R}$-linear. Therefore this is an $\mathbb{R}$-algebra. (Note that in Exercise 1.4.3 you have proven that $(\mathbb{H}, \cdot)$ is indeed associative, moreover $\mathbb{H}^* = \mathbb{H} \setminus 0$ is a group.) Because all elements (except for 0) are invertible, this is called a *division algebra (可分代数)*.

4. Remember that the structure of $\mathbb{H}$ essentially arises from that of $Q := \langle \mathsf{i}, \mathsf{j} : \mathsf{i}^4 = 1, \mathsf{i}^2 = \mathsf{j}^2, \mathsf{iji}^{-1} = \mathsf{j}^{-1} \rangle$. The generalization is the following: Given a (discrete / finite) (semi)-group G together with a field F, then the *group algebra (群的代数)* $F[G]$ is the vector space $\langle \delta_g : g \in G \rangle_F$ together with the F-bilinear operation $*\colon F[G] \times F[G] \to F[G] : \delta_g * \delta_h = \delta_{gh}$ for all $g, h \in G$ (and F-linearly extended).

5. Given two algebras A and B over the same field F, we can consider their *tensor product (张量积)* $A \otimes B = \langle a \otimes b : a \in A, b \in B \rangle_F$ and define a multiplication as $(a_1 \otimes b_1)(a_2 \otimes b_2) := (a_1 a_2) \otimes (b_1 b_2)$ and F-linear extension. An example of that is the tensor product of (square) matrices which obeys the law $\mathrm{End}(F^m) \otimes \mathrm{End}(F^n) \cong \mathrm{End}(F^{mn})$ as induced by the tensor product of the underlying vector spaces $F^m \otimes F^n \cong F^{mn}$.

6. Consider the real vector space $\mathbb{R}^3$ together with the vector product $\times\colon \mathbb{R}^3 \otimes \mathbb{R}^3 \to \mathbb{R}^3$. This operation is also bilinear and therefore $(\mathbb{R}^3, \times)$ is an algebra over $\mathbb{R}$. Note however that this algebra is not associative, because $(a \times b) \times c \neq a \times (b \times c)$ for all $a, b, c \in \mathbb{R}^3$. It fulfills however another nice property which makes it a Lie algebra (李代数).

Note that many of the notions introduced in the further sections have analogs for algebras.

2.1.99 Exercises

Exercise 2.1.1. Let $(A, +, \cdot)$ be any algebra with neutral element 0 (w.r.t. addition). Show that $0a = 0 = a0$ for every $a \in A$.

Exercise 2.1.2. Let $(R, +, \cdot)$ be a set with two monoidal operations $+$ and $\cdot$ neither of which need to be commutative, but both are associative, $\cdot$ is distributive over $+$, i.e.

$$(a + b)c = ac + bc,$$
$$a(b + c) = ab + ac,$$

and $+$ has inverse elements $-a \in R$ for every $a \in R$. Show that $(R, +, \cdot)$ is a (non-necessarily commutative) ring, i.e. $+$ is abelian.
Hint: Consider products $(a + b)(c + d)$.

Exercise 2.1.3. a. Given an abelian group $(A, +)$. Show that its group-endomorphisms $\mathrm{End}(A, +)$ form a unital non-commutative associative ring.

b. Given any ring $(R, +, \cdot)$. Show that $(R, +, \cdot)$ embeds canonically into $\mathrm{End}(R, +)$.

Exercise 2.1.4. a. Given the semi-group algebra $F[G] = \langle \delta_g : g \in G \rangle_F$ with the multiplication $\delta_g * \delta_h = \delta_{gh}$ show that this is associative as long as G is a semi-group.

b. Given the polynomial multiplication as defined in the lecture, i.e.

$$(a_0 + a_1 x + \cdots + a_m x^m)(b_0 + b_1 x + \cdots + b_n x^n) := \sum_{k=0}^{m+n} \sum_{i+j=k} a_i b_j x^k$$

show that it is associative.

Exercise 2.1.5. Let $\alpha \in \mathbb{C}$ be a zero of a non-trivial polynomial over $\mathbb{Z}$. Show that $\mathbb{Z}[\alpha] = \langle 1, \alpha, \alpha^2, \ldots \rangle_{\mathbb{Z}}$ is a ring together with an embedding $e \colon \mathbb{Z} \to \mathbb{Z}[\alpha] : n \mapsto n \cdot 1$.

a. Show that the Gaussian[2] integers $\mathbb{Z}[\mathsf{i}] = \{a + b\mathsf{i} : a, b \in \mathbb{Z}\}$ form a ring. Find its units, i.e. those elements $u \in \mathbb{Z}[\mathsf{i}]$ that have a multiplicative inverse $v \in \mathbb{Z}[\mathsf{i}]$, i.e. $uv = 1 = vu$. Show that these elements form a group (under multiplication).

b. Show that also $\mathbb{Z}[\sqrt{-2}] \subset \mathbb{C}$ form a subring of the complex numbers. Find its units.

c. What happens when $\alpha \in \mathbb{C}$ is transcendental over $\mathbb{Q}$, i.e. not root of any non-trivial polynomial?

Exercise 2.1.6. Let R be a commutative ring (not necessarily with unit). Show that $R^1 := \mathbb{Z} \ltimes R$ with operations $(m, a) + (n, b) := (m+n, a+b)$ and $(m, a)(n, b) := (mn, ab + mb + na)$ is a unital ring. What is its multiplicative identity?

[2] J. Carl Friedrich Gauss *1777/4 in Brunswick, Germany †1855/2

2.2 Homomorphisms, subrings, and ideals (理想)

Corresponding to groups, the next important notion is that of a ring homomorphism:

Definition 2.2.1. *Given two rings R and S, a* ring-homomorphism (环同态) *is a map $\phi\colon R \to S$ such that ϕ is a homomorphism of the additive groups, preserves multiplication and maps $1 \in R$ to $1 \in S$, i.e.*

$$\phi(a+b) = \phi(a) + \phi(b), \tag{2.3}$$
$$\phi(ab) = \phi(a)\phi(b), \tag{2.4}$$
$$\phi(1) = 1. \tag{2.5}$$

We say that the ring homomorphism ϕ is injective (monomorphism, embedding) *iff* $\ker\phi := \{r \in R : \phi(r) = 0\} = \{0\}$. *We say that ϕ is* surjective (epimorphism) *iff* $\operatorname{im}\phi := \phi(R) := \{\phi(r) : r \in R\} = S$. *We say that ϕ is an* isomorphism *iff it is injective and surjective. In the latter case we also say that R is isomorphic to S ($R \cong S$ via ϕ).*

Example 2.2.2. 1. Consider the map $e\colon \mathbb{Z} \to \mathbb{Q} : z \mapsto {}^{z}/_{1}$. Clearly this is a ring homomorphism. Note that its kernel is $\ker e = \{0\}$, thus it is an embedding and so $\mathbb{Z}$ operates inside $\mathbb{Q}$.

Analogous to groups, we are also lead to the following two notions:

Definition 2.2.3. *Given a (unital) ring R, then a* subring (环子) *is a subset $S \subset R$ that forms a (unital) ring under the same operations.*

An ideal (理想) *$I \triangleleft R$ is an additive subgroup $I \subset (R, +)$ such that $RI \subset I$.*

Example 2.2.4. 0. The trivial ideals are 0 and R. The latter also is the trivial subring. The smallest possible subring is $\langle 1\rangle \subset R$, i.e. the image of the following ring homomorphism $\phi_0\colon \mathbb{Z} \to R : n \mapsto n \cdot 1$.

1. It also follows that $\ker\phi \triangleleft R$ is an ideal and $\operatorname{im}\phi \subset S$ is a subring for every ring homomorphism $\phi\colon R \to S$.

2. Note further that $1 \in I$ implies $I \supset RI = R$, i.e. the ideal automatically is trivial.

3. $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ are subrings.

4. Starting from a field F, we see that for every nonzero ideal $(0) \neq I \triangleleft F$ there is a non-zero element $i \in I$ and thus $FI \supset Fi = F$, i.e. the ideal is already the whole field F. Thus fields only have trivial ideals. In particular the only homomorphisms between fields are embeddings (because $\phi(1) = 1 \neq 0$ implies $\ker\phi \neq F$).

5. Note that the $(n) := nR$ for $n \in R$ are the *principal ideals (主理想)* of R. In particular the Euclidean algorithm shows that all ideals of $\mathbb{Z}$ are of this form ($\{0\} = (0)$).

6. Given a family of ideals $\{I_\alpha : \alpha \in A\}$ then their intersection $\bigcap_{\alpha\in A} I_\alpha$ is an ideal (analogously to normal subgroups). We call an ideal I *irreducible (不可约理想)* iff it cannot be written as the intersection of two different ideals.

7. Given a ring $(R, +, \cdot)$ we can construct further ideals in the following way. Let $\{a_j \in R : j \in J\}$ be ring elements and consider the smallest ideal that contains all elements a_j, i.e. $\bigcup_{\{a_j\}\subset I \lhd R} I \lhd R$. If there are only finitely many elements, we denote their *generated ideal (生成理想)* as $(a_1, \ldots, a_n)$.

8. Given two ideals $I, J \lhd R$, then their sum $I + J$ is another ideal, because $(I+J)R = IR+JR \subset I+J$. In this way the ideal generated by the elements a_j for $j \in J$ is $\sum_{j\in J}(a_j) \lhd R$, i.e. the sum of principal ideals. Endowed with these two operations the ideals of a ring form a *distributive lattice (分配格)* where we can also write $\vee$ for $+$, because $I + J$ is the ideal generated by $I \cup J$. You should show the distributive law $(I \cap (J + K) = I \cap J + I \cap K$ or equivalently $(I + J) \cap K = I \cap K + J \cap K$ for all ideals $I, J, K \lhd R$) as a homework.

9. Given two ideals $I, J \lhd R$, then their product is defined as $IJ := \langle ij : i \in I, j \in J\rangle_R$. Note that this also contains finite sums of products of elements. Clearly $RIJ \subset IJ$ and $IJ \subset I \cap J$, but in general it is not the same, e.g. $I = J = (x + 1) \lhd \mathbb{R}[x]$, then $I^2 = (x + 1)^2\mathbb{R}[x] \subsetneq (x + 1)\mathbb{R}[x]$. Also this product is distributive over the addition, i.e. $I(J + K) = IJ + IK$ for ideals $I, J, K \lhd R$.

10. Given a ring R. Its *nilradical (诣零根)* is the set $\operatorname{nil\,rad} R := \sqrt{(0)} := \{z \in R : z^n = 0 \text{ for some } n \in \mathbb{N}\}$. In the homework you will show that this is an ideal.

Remark 2.2.5. Note that in the case of non-commutative (associative) algebras, we need to distinguish between left ideals ($RI \subset I$), right ideals ($IR \subset I$) and two-sided ideals (both). The kernel of an algebra homomorphism is both. Also for the following construction we need a two-sided ideal.

Another construction that follows immediately is that of a quotient ring by an ideal:

Proposition 2.2.6 (Factor ring)**.** *Given a non-trivial ideal $I \lhd R$, then the space of cosets R/I has an induced ring structure via representatives $(a + I) + (b + I) =$*

$(a+b)+I$, $(a+I)(b+I) = ab+I$ with $0+I = I$ and $1+I$ being the neutral elements. Moreover the projection $\pi\colon R \to R/I$ is a surjective ring homomorphism. This is denoted the quotient ring (商环).

Proof. Since $RI \subset I$ and for $1 \in R$, we see that the multiplication is indeed the operation on the sets $ab+I = \{(a+i)(b+j)+k : i, j, k \in I\}$ and thus representative independent. Also for $1 \notin I$ it is clear that $1 + I \neq I$, i.e. the quotient ring is indeed unital. □

Example 2.2.7. Given any principal ideal $(n) \triangleleft R$, we obtain a quotient ring $R/(n)$. Particular examples are the integers modulo n where $n \in \mathbb{Z}$.

Lemma 2.2.8. *Given a ring homomorphism $\phi\colon R \to S$ and an ideal $I \triangleleft R$, then ϕ factors through the projection $\pi\colon R \to R/I$,i.e. $\phi = \bar{\phi} \circ \pi$ for some homomorphism $\bar{\phi}\colon R/I \to S$, iff $\ker \phi \subset I$. In this case ϕ factors uniquely.* □

Also analogous to groups we have the standard isomorphism theorems:

Theorem 2.2.9 (First isomorphism theorem)**.** *Given a ring homomorphism $\phi\colon R \to S$, then $R/\ker \phi \cong \operatorname{im} \phi$.* □

Theorem 2.2.10 (Second isomorphism theorem)**.** *Given two ideals $I, J \triangleleft R$ with $I \subset J$, then ($I \triangleleft J$ is an ideal in the non-unital ring J), $J/I \triangleleft R/I$ is an ideal and*

$$(R/I)/(J/I) \cong R/J.$$

The proof is left as an exercise.

Theorem 2.2.11 (Third isomorphism theorem)**.** *Given a subring $S \subset R$ and an ideal $I \triangleleft R$, then $S + I \subset R$ is a subring, $S \cap I \triangleleft S$ is an ideal, and*

$$(S + I)/I \cong S/(S \cap I)$$

Also this proof is left as an exercise.

Proposition 2.2.12 (Characteristic)**.** *Given any (commutative unital) ring $(R, +, \cdot, 1)$ there is a canonical ring homomorphism $\phi_0\colon \mathbb{Z} \to R : n \mapsto n \cdot 1$. Its kernel is an ideal in $\mathbb{Z}$ and thus generated by a unique non-negative element $n \in \mathbb{N}$. We call this $n = \operatorname{char} R$ the* characteristic (特征) *of R. In particular ϕ_0 factors to $\bar{\phi}\colon \mathbb{Z}/(n) \hookrightarrow R$.* □

2.2.99 Exercises

Exercise 2.2.1. a. Show that the union $\bigcup_{n\geq 0} S_n$ of an ascending chain of subrings $\langle 1\rangle \subset S_1 \subset S_2 \subset \cdots \subset R$ is a subring of R.

b. What happens for ascending chains of ideals?

c. Show that every intersection of subrings $S_\alpha \subset R$, $\alpha \in A$ is a subring $\bigcap_{\alpha\in A} S_\alpha \subset R$.

d. Show the corresponding fact for ideals.

Exercise 2.2.2. Let $I, J \lhd R$ be ideals. Show that $I \cup J$ is an ideal iff $I \subset J$ or $J \subset I$.

Exercise 2.2.3. Given any ring R. Show that the nil radical $\operatorname{nil\,rad} R := \{z \in R : z^n = 0 \text{ for some } n \in \mathbb{N}\}$ is an ideal. What is the nil radical of $\mathbb{Z}$?

Exercise 2.2.4. Prove the isomorphism theorems for rings (Theorem 2.2.9, 2.2.10, and 2.2.11, respectively).

Exercise 2.2.5. Given any ring homomorphism $\phi\colon R \to R'$. Show the following

a. a chain of subrings $S_1 \subset S_2 \subset R$ corresponds to a chain of subrings $S_1' \subset S_2' \subset R'$ where $S' := \phi(S) \subset R'$;

b. any chain of subrings $S_1' \subset S_2' \subset R'$ corresponds to a chain of subrings $\ker\phi \subset S_1 \subset S_2 \subset R$ where $S := \phi^{-1}(S') := \{r \in R : \phi(r) \in S'\}$;

c. any chain of ideals $I_1' \subset I_2'$ with $I_k' \lhd R'$ corresponds to a chain of ideals $\ker\phi \subset I_1 \subset I_2$ with $I_k := \phi^{-1}(I) \lhd R$;

d. ϕ and ϕ^{-1} also preserve intersections (of subrings or ideals) and sums and products of ideals.

Exercise 2.2.6. Let $\phi\colon R \to R'$ be a ring homomorphism and $I \lhd R$ an ideal.

a. Assuming that ϕ is surjective, show that $\phi(I) \lhd R'$ is an ideal.

b. Given the results in Exercise 2.2.5 and a surjective ϕ, show that there is a 1:1 correspondence between ideals $\ker\phi \subset I_1 \subset I_2$ with $I_k \lhd R$ and $I_1' \subset I_2'$ with $I_k' \lhd R'$.

Hint: You also have to show uniqueness of $\ker\phi \subset I \lhd R$ with $\phi(I) = I'$ for any fixed $I' \lhd R'$.

c. Give an example where $\phi(I)$ is not an ideal.

2.3 Domains (整环) and fields (域)

Rings were modeled after the integers. Note however that the integers have more particular properties than arbitrary rings. One is that the product of two non-zero integers is always nonzero. As we have seen from the example of $\mathbb{Z}/(6)$ this is generally not true for rings. Instead we define:

Definition 2.3.1. *1. A (commutative) ring (with unit 1) is called* (integral) domain (整环) *if there are no zero-divisors.*

2. Analogously to integers we introduce the notion a divides b, $a|b$ if there exists an $q \in R$ such that $b = qa$.

3. R is called a field (域) *iff the multiplication $(R \setminus \{0\}, \cdot)$ forms a group.*

Example 2.3.2. 0. As mentioned before the integers are a domain.

1. The integers $\mathbb{Z}/(n)$ modulo a positive integer $n > 0$ form a domain iff $n \in \mathbb{P}$ is a prime. In this case $\mathbb{F}_p := \mathbb{Z}/(n)$ is even a field.

2. Note that for polynomials $p, q \in R[x]$ and a domain R (e.g. a field), then the degree obeys the formula $\deg(p \cdot q) = \deg p + \deg q$. Therefore the polynomials with coefficients in a domain form a domain themselves.

3. Given any ring R, then the quotient by a *maximal ideal (极大理想)* $\mathfrak{m} \triangleleft R$ (maximality with respect to inclusion and $\mathfrak{m} \neq R$) is a field, because every element of the quotient that has no multiplicative inverse gives rise to a bigger nontrivial ideal.

4. Given a ring R, then an ideal $\mathfrak{p} \triangleleft R$ is called *prime ideal (素理想)* if the quotient ring $R/\mathfrak{p}$ is a domain. A ring is thus a domain iff (0) is a prime ideal.

Proposition 2.3.3. *Given a domain R, we can mimic the construction of the rational numbers to obtain a field. Let $K[R] := (R \times (R \setminus \{0\}))/\sim$ where we write the elements of $K[R]$ as a/b (with $b \neq 0$) and say $a/b = c/d$ iff $ad = bc$. Addition of two fractions is done via expansion to a common denominator, i.e.*

$$a/b + c/d = (ad + bc)/(bd),$$

the negative of a/b is $(-a)/b$, and multiplication is done component-wise, i.e.

$$(a/b) \cdot (c/d) = (ac)/(bd)$$

with the unit $1/1$ and the inverse to a/b is b/a for $a/b \neq 0$. This is called the field of fractions (分式环).

Proof. Since R is a domain it is not hard to check that $K[R]$ is indeed a field. □

Note that $K[R]$ fulfills the *universal property (泛性质)* that for every injective ring homomorphism $\phi\colon R \to F$ into a field, there is a unique extension $\tilde{\phi}\colon K[R] \to F$.

Definition 2.3.4. *Given a ring R, then we define the following notions:*

1. *A* unit (单位) *is an element $u \in R$ that has a multiplicative inverse, i.e. $\exists u^{-1} \in R : uu^{-1} = 1$. The set of units is denoted by R^* and forms a group.*

2. *An ideal $I \triangleleft R$ is called* irreducible (不可约理想) *if it cannot be written as the intersection of strictly larger ideals.*

3. *Given an integral domain, a non-zero non-unit element $p \in R$ is called* irreducible (不可约元素) *if in every decomposition as $p = ab$ with $a, b \in R$ either a or b is a unit.*

Example 2.3.5. 0. Given the integers $\mathbb{Z}$, then $\mathbb{Z}^* = \{\pm 1\}$. The irreducible ideals are the ideals generated by a prime $p \in \mathbb{P} = \{2, 3, 5, 7, 11, \dots\}$ and the primes (together with their negatives) are the irreducible elements.

1. Consider the ring $\mathbb{Z}[x]$ of polynomials with coefficients in integers. Note that (x) is a prime ideal, because $\mathbb{Z}[x]/(x) \cong \mathbb{Z}$. It is not maximal, because $\mathbb{Z}$ is not a field. More particularly $0 \neq n \in \mathbb{Z}$ implies that $(x) \subsetneq (x, n) \triangleleft \mathbb{Z}[x]$ is a larger ideal containing (x).

2. $(x^n) \triangleleft \mathbb{Q}[x]$ is irreducible, because $I \cap J = (x^n)$ implies $f \equiv 0 \pmod{x^n}$ for all $f \in I, J$ and also I or $J \not\equiv 0 \pmod{x^{n+1}}$. But then one of them has to be (x^n). However (x^n) is a prime ideal only for $n = 1$ ($x + (x^n) \in R[x]/(x^n)$ is nilpontent).

3. You can find an example of an irreducible non-prime element in the counter examples of unique factorization domains (Example 2.5.3-4).

The following conclusions however are true.

Proposition 2.3.6. *Given any ring R, then*

1. *every maximal ideal $\mathfrak{m} \triangleleft R$ is prime,*

2. *every prime ideal $\mathfrak{p} \triangleleft R$ is irreducible,*

3. *every* prime element (质) *$p \in R$, i.e. p non-zero non-unital and $p|ab$ for $a, b \in R$ implies $p|a$ or $p|b$, is irreducible,*

4. $(p) \triangleleft R$ *for* $p \in R \setminus 0$ *is a prime ideal iff* $p \in R$ *is prime.*

Proof.

1. a field e.g. $R/\mathfrak{m}$ is in particular a domain.

2. if $\mathfrak{p} \triangleleft R$ is not irreducible, i.e. $\mathfrak{p} = I \cap J$ for $\mathfrak{p} \subsetneq I, J$, then in particular there are $i \in I$ and $j \in J$ with $i, j \notin \mathfrak{p}$ but $ij \in IJ \subset I \cap J = \mathfrak{p}$. By Exercise 2.3.1 $\mathfrak{p}$ is then not prime.

3. if p were reducible, i.e. $p = ab$ with $a, b \in R$ both not units, then $p|ab$, but $p \nmid a$ and $p \nmid b$.

4. Note that the notations $a \equiv 0 \pmod{(p)}$ and $p|a$ are equivalent for $a, p \in R$.

□

2.3.1 Properties of polynomial rings

Remember the definition of degree of a polynomial $p \in R[x]$ over an arbitrary ring R in an indeterminate x, i.e. $\deg_x p = n$ if $p = a_0 + a_1 x + \cdots + a n_x^n$ with $a_i \in R$ and $a_n \neq 0$. We have the following properties:

Proposition 2.3.7. *For* $p, q \in R[x]$ *with* $p, q \neq 0$ *the zero polynomial, then*

$$\begin{aligned} \deg(p+q) &\leq \max(\deg p, \deg q), \quad \textit{and ``}= \textit{ if } \deg p \neq \deg q, \\ \deg(pq) &\leq (\deg p) + (\deg q), \quad \textit{and ``}= \textit{'' if } R \textit{ is a domain.} \end{aligned}$$

Remark 2.3.8. Remember the long polynomial division, i.e. given polynomials $p, d \in F[x]$ where F is any field and $d \neq 0$, then there are unique polynomials $q, r \in F[x]$ such that $p = qd + r$ and $\deg r < \deg d$. The explicit computation is

$$\begin{array}{rrrrl} (x^3 & +2x^2 & +3x & +1 &) : (x+1) = x^2 + x + 2 \text{ rem } -1 \\ -(x^3 & +x^2) & & & \\ \hline & x^2 & +3x & & \\ & -(x^2 & +x) & & \\ \hline & & 2x & +1 & \\ & & -(2x & +2) & \\ \hline & & & -1 & \end{array}$$

By inspection this algorithm also succeeds if R is any ring and $\operatorname{lc} d := d_n \in R^*$ is a unit. Uniqueness is however only guaranteed if R is a domain.

Note also that polynomials can be used as functions, i.e. there is the evaluation map $\mathrm{ev}\colon R[x] \times R \to R : (a_0 + a_1 x + \cdots + a_n x^n, b) \mapsto a_0 + a_1 b + \cdots + a_n b^n$ with the following property.

Proposition 2.3.9. *For a ring R and any element $a \in R$, the evaluation map $\mathrm{ev}_a\colon R[x] \to R : p \mapsto \mathrm{ev}(p, a)$ is a ring homomorphism.*

Proof. This uses the commutativity of R, and then the fact that the polynomial multiplication of x forms a free monoid. □

Remark 2.3.10. Note that for arbitrary rings even $\mathrm{ev}_a(p) = \mathrm{ev}_a(q)$ for all $a \in R$ and fixed $p, q \in R[x]$ does not imply $p = q$, not even over fields. Consider for example $x^p, x \in \mathbb{F}_p[x]$. Due to Fermat's[3] little theorem $a^{p-1} \equiv 1 \pmod p$ for $a \in \mathbb{Z}$, $p \not| a$ and thus $\mathrm{ev}(x^p, a) = \mathrm{ev}(x, p)$ for every $a \in \mathbb{F}_p \cong \mathbb{Z}/(p)$. But certainly $x^p \neq x \in \mathbb{F}_p[x]$ in the way we defined polynomials.

Remember the meaning of zeros of a polynomial:

Proposition 2.3.11. *Given any non-zero polynomial $p \in R[x]$ over a domain R. We call $r \in R$ a* root (根) *of p if $\mathrm{ev}(p, r) = 0$. In this case $x - r$ divides p.*

Proof. Assume that $p = (x - r)q + r'$. Since $\deg r' < \deg(x - r) = 1$ we have $r' \in R$. But $0 = \mathrm{ev}_r(p) = \mathrm{ev}_r((x - r)q + r') = \mathrm{ev}_r(x - r)\,\mathrm{ev}_r(q) + r' = r'$ shows that $r' = 0$, i.e. the division has no remainder. □

Therefore every non-zero polynomial of degree d over a domain has at most d roots (even when counted with multiplicity). Conversely we see that a domain that has a root for every polynomial (called *algebraically closed (代数闭的)*) must be at least a field, because all the $rx - 1 \in R[x]$ for $r \in R \setminus 0$ have a root. However the example of $x^2 + 1 \in \mathbb{Q}[x]$ shows that being a field is not sufficient.

Note that in general it is important to keep track of the base ring for a polynomial, i.e. $x^2 + 1 \in \mathbb{F}_2[x]$ and $x^2 + 1 \in \mathbb{Q}[x]$ are not the same polynomial. Given a ring homomorphism we are however able to map the polynomial rings as well.

Proposition 2.3.12. *Every ring homomorphism $\phi\colon R \to S$ induces a homomorphism of the associated polynomial rings $\phi\colon R[x] \to S[y] : (a_0 + a_1 x + \cdots + a_n x^n) \mapsto (\phi(a_0) + \phi(a_1) y + \cdots + \phi(a_n) y^n)$ that commutes with evaluation as $\mathrm{ev}^y(\phi(p), \phi(a)) = \phi(\mathrm{ev}^x(p, a))$ for all $p \in R[x]$ and $a \in R$.* □

Conversely the homomorphisms from $R[x]$ into any (associative) algebra consist of a base ring homomorphism together with an evaluation as follows:

[3] Pierre de Fermat *1601/8 or 1607/8 in Beaumont-de-Lomagne/France, †1665/1

Proposition 2.3.13. *Given a ring R and an (associative unital) algebra A, then every homomorphism $\phi\colon R[x] \to A$ is $\phi = \widetilde{\mathrm{ev}}_a \circ \phi_0$ for $\phi_0\colon R \to S := \phi(R) \subset A$, $a = \phi(x) \in \mathrm{cent}_A(S)$ and $\widetilde{\mathrm{ev}}_\bullet$ the restriction of the evaluation map on $A[x] \times A$ to $S[x] \times \mathrm{cent}_A(S) \to \mathrm{cent}_A(S)$ which is a homomorphism in the argument in $S[x]$. In particular every pair (ϕ_0, a) with $\phi_0\colon R \twoheadrightarrow S \subset A$ and $a \in \mathrm{cent}_A(S)$ is in 1:1 correspondence with a homomorphism ϕ.*

Proof. Note that ϕ is obviously determined by the individual images in $S := \phi(R)$ of each element in R and $x \in R[x]$. If ϕ is a ring homomorphism, then $\phi(x)$ must commute with every $s \in S$, thus $a := \phi(x) \in \mathrm{cent}_A(S)$ for the subring $S \subset A$. □

Remark 2.3.14. It is also possible to extend the notion of polynomials to (associative non-commutative) algebras, either in the naive way by considering finite sequences together with the component-wise addition and multiplication laws of commutative polynomials. Note however that in this case the evaluation map ev_a is only an algebra homomorphism for $a \in \mathrm{cent}\, A$.

The alternative is to consider the tensor algebra $T'(A) := \bigoplus_{n\geq 1} A^{\otimes n}$ of the vector space A/F together with the component-wise addition and the multiplication

$$(a_0 \otimes a_1 \otimes \cdots \otimes a_m)(b_0 \otimes \cdots \otimes b_n) := (a_0 \otimes \cdots \otimes a_m b_0 \otimes b_1 \otimes \cdots \otimes b_n)$$

for $a_i, b_i \in A$. In this case the evaluation map is

$$ev_b(a_0 \otimes \cdots \otimes a_n) := a_0 b a_1 b \ldots b a_n.$$

For every $b \in A$. Then $\mathrm{ev}_b\colon T'(A) \to A$ is an algebra homomorphism. And if A has no zero-divisors, then $\deg(a_0 \otimes \cdots \otimes a_n) := n$ for $a_i \neq 0$ is the degree in the sense that $\deg(ab) = (\deg a) + (\deg b)$ for all $a, b \in T'(A)$.

Remember the properties of derivatives, i.e. $(f(x) + g(x))' = f'(x) + g'(x)$, $(cf(x))' = cf'(x)$ and $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$. These generalize to arbitrary rings as follows.

Definition 2.3.15. *A differential algebra is an algebra A together with an additive map $\partial\colon A \to A$ that fulfills the Leibniz rule*

$$\partial(fg) = (\partial f)g + f(\partial g) \quad \forall f, g \in A.$$

We call $\mathrm{Const}(A, \partial) := \ker \partial$ *the constants.*

Example 2.3.16. 0. The integers and their quotient rings as well as the rational numbers (field of fractions) only have trivial derivatives $\partial = 0$, i.e. are constant.

1. The polynomial ring $R[x]$ has plenty of derivatives, e.g. $\partial\colon c \mapsto 0, x \mapsto p$ for every $c \in R$, any fixed $p \in R[x]$, and extended using the Leibniz rule and additivity.

Proposition 2.3.17. *Given a differential ring $(R, 1, \partial)$, then $\phi\colon \mathbb{Z} \to R : n \mapsto n \cdot 1$ has $\operatorname{im}\phi \subset \operatorname{Const}(R, \partial)$. Conversely for every $c \in \operatorname{Const}(R, \partial)$ we have*

$$\partial(cf) = c\partial f \quad \forall f \in R$$

Proof. Start from the unit $\partial 1 = \partial(1^2) = 2\partial 1$ which leads to $\partial 1 = 0$. By additivity $\partial(n \cdot 1) = 0$ for every $n \in \mathbb{Z}$, i.e. $\operatorname{im}\phi \subset \operatorname{Const}(R, \partial)$. The general Leibniz rule then implies the rule for multiplication with a constant. □

2.3.2 Polynomials in several indeterminates

It is also possible to define polynomials in several variables, e.g. as follows:

Definition 2.3.18. *Let R be any ring and $x_1, x_2, \ldots$ countably many indeterminates. We define the polynomial ring $R[x_1, x_2, \ldots]$ as the inductive limit $\bigcup_{n\geq 0} R[x_1][x_2]\ldots[x_n]$.*

Lemma 2.3.19. *If there are only finitely many indeterminates, we have $R[x_1, \ldots, x_n] = R[x_1][x_2]\ldots[x_n]$ and in particular $R[x][y] = R[y][x]$.* □

Proposition 2.3.20. *Given a polynomial ring in several variables $R' := R[x_1, x_2, \ldots]$ beside the partial degrees $\deg_i\colon R' \to \mathbb{N} \cup \{-\infty\}$ with $\deg_i(x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}) := n_i$ we also have the* total degree $\deg\colon R' \to \mathbb{N} \cup \{-\infty\}$ *with*

$$\deg(x_1^{n_1} x_2^{n_2} \ldots x_k^{n_k}) := n_1 + n_2 + \cdots + n_k.$$

Every degree obeys the rules

$$\deg_\alpha(p+q) \leq \max(\deg_\alpha p, \deg_\alpha q), \quad \text{and "=" for } \deg_\alpha p \neq \deg_\alpha q, \tag{2.6}$$

$$\deg_\alpha(pq) \leq (\deg_\alpha p)(\deg_\alpha q), \quad \text{and "=" for } R \text{ a domain}, \tag{2.7}$$

$$\deg p \leq \sum_i \deg_i p, \quad \text{and "=" for monomials.} \tag{2.8}$$

Idea of proof. For the partial degrees that follows from the notation $R[x_1, x_2, \ldots] = R[x_1, x_2, \ldots, \hat{x}_i, \ldots][x_i]$ and the corresponding fact for polynomials in the indeterminate x_i. For the total degree the proof is analogous to that in one variable, just a bit longer. The last inequality follows from the previous ones and the fact that $\deg = \sum_i \deg_i$ for monomials. □

Analogously to the case of one indeterminate, we also have (partial) evaluations here, i.e. maps $ev^k\colon R[x_1, x_2, \ldots] \times R \to R[x_1, x_2, \ldots, \hat{x}_k, \ldots]$ by replacing the kth indeterminate by an element $a \in R$. Analogous to the case of one indeterminate we obtain.

Corollary 2.3.21. *Given a ring R and an element $a \in R$, then the kth evaluation $ev_a^k \colon R[x_1, x_2, \dots] \to R[x_1, x_2, \dots, \hat{x}_k, \dots]$ is a ring-homomorphism.*

Remark 2.3.22 (Substitution)**.** Since $R[x]$ is a ring in itself, it is also possible to define substitutions as a special case of evaluations in a ring with several variables. Let thus $R' := R[x]$ with the obvious homomorphism $\mathrm{subs}_{x\to y} \colon R[x] \to R'[y] : R \xrightarrow{\sim} R, x \mapsto y$. Then we define the general substitution $\mathrm{subs}_{x\to t} \colon R[x] \to R[x] : p \mapsto \mathrm{ev}_t^2(\mathrm{subs}_{x\to y}(p)) \in R[x]$ for arbitrary $t \in R[x]$. You can show as a (rather formal) exercise that this also is a ring-homomorphism. Moreover it is possible to extend this to several variables (and simultaneous substitutions).

Once we have several variables, it is also possible to define partial derivatives when we remember Schwarz' theorem: $\partial_j \partial_i = \partial_i \partial_j$

Definition 2.3.23. *A partial differential ring is a (commutative) ring $(R, +, \cdot)$ together with commuting derivatives $\partial_1, \partial_2, \dots$.*

Example 2.3.24. In the ring $R[x_1, x_2, \dots]$ it is possible to introduce the partial derivatives $\partial_i \colon R \to 0, x_j \mapsto \delta_{ij}$ which is 1 for $i = j$ and 0 otherwise.

2.3.3 Formal power series

Sometimes it is necessary to consider formal limits of infinite sequences of polynomials, i.e. consider the truncation p_k of the infinite Taylor series after k elements and now take the formal (projective) limit $k \to \infty$. The result may not be a function (e.g. when the original function was not analytic), but it can still be considered as an infinite sequence and manipulated analogously to (finite) polynomials:

Definition 2.3.25. *Given a ring R, the formal power series $R[[x]]$ in one indeterminate x over R are the infinite sequences $(a_n)_{n=1}^{\infty}$ with $a_n \in R$, component-wise addition and the multiplication rule*

$$(a_0 + a_1 x + a_2 x^2 + \dots)(b_0 + b_1 x + b_2 x^2 + \dots) := a_0 b_0 + \sum_{n\geq 1} \sum_{i+j=n} a_i b_j x^n.$$

Note that for finite n this contains only finitely many terms and thus we end up with coefficients in R. It is easy to see that this is a ring and moreover a domain if R is a domain.

Example 2.3.26. Consider the Taylor series of the (analytic) exponential function $\exp x = 1 + \sum_{n\geq 1} \frac{1}{n!} x^n$ These multiply as

$$\exp(x)\exp(y) = \sum_{j,k\geq 0} \frac{1}{j!k!} x^j y^k = \sum_{n\geq 0} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \sum_{n\geq 0} \frac{1}{n!} (x+y)^n = \exp(x+y)$$

Thus the exponential map is a homomorphism from $(F,+)$ to $(F^* + xF[[x]],\cdot)$ the formal power series in one indeterminate with nonzero absolute term. It is part of an exercise to see what the inverse elements on the right hand side are and thus to make it into a group.

Unfortunately the formal power series have less (obvious) homomorphisms to the base ring, i.e. the only well-defined evaluation is evaluation at 0, i.e. $\mathrm{ev}_0\colon R[[x]] \to R : (a_n) \mapsto a_0$ which is obviously (still) a ring homomorphism.

It is however still possible to use the previously defined derivatives, e.g. for any $t \in R[[x]]$, $\partial\colon R[[x]] \to R[[x]] : R \mapsto 0, x \mapsto t$ extends uniquely to a derivation.

2.3.99 Exercises

Exercise 2.3.1. Show that an ideal $\mathfrak{p} \lhd R$ in any ring R is a prime ideal iff $xy \in \mathfrak{p}$ for $x, y \in R$ implies $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$.

Exercise 2.3.2. Let $(S,\cdot)$ be an abelian monoid (commutative semigroup with neutral element id) that is cancellative, i.e. for every $a, b, c \in S$, $ab = ac$ implies $b = c$. Construct a group of fractions $K[S]$ and state and show its universal property.
Hint: The universal property should consider maps into any abelian group $(A,\cdot)$.

Exercise* 2.3.3. Given a non-commutative (unital) integral ring R (i.e. an associative $\mathbb{Z}$-algebra with $ab = 0$ implies $a = 0$ or $b = 0$) that fulfills the Ore condition: Every finite intersection of non-trivial principal ideals is nontrivial. Show that the analogon of the field construction gives a division algebra, i.e. $S[R] := (R \times R \backslash 0)/\!\sim$ where $a/b \sim c/d$ iff $ad = cb$ (in that order). Show that

0. $\sim$ is an equivalence relation;

a. the addition $a/b + c/d = (af + bg)/p$ for $a, b, c, d \in R$, $c, d \neq 0$ and $p, f, g \in R \setminus 0$ such that $bf = p = dg$ is well-defined and forms an abelian group. Note that you have to show existence of some (p, f, g) as well as $a/b + c/d \sim a'/b' + c'/d'$ for all pairs $a/b \sim a'/b'$ and $c/d \sim c'/d'$. (What is the neural element, the inverses?)

b. the multiplication $(a/b) * (c/d) = \tilde{a}/\tilde{d}$ for $a, b, c, d \in R$, $b, c, d \neq 0$, and some $\tilde{a}, \tilde{b}, \tilde{d} \in R$ with $\tilde{a}/\tilde{b} \sim a/b$ and $\tilde{b}/\tilde{d} \sim c/d$. Extend by $(a/b) * (0/d) = (0/1)$ and show that multiplication is also well-defined and gives a (non-commutative) ring structure.

c. Show that $S[R]$ is a division-ring generated by $\imath\colon R \to S[R] : a \mapsto a/1$. You can, e.g. show that $(a/b)/(c/d) = \tilde{a}/\tilde{c}$ for $a, b, c, d \in R$ with $b, c, d \neq 0$ and

some $\tilde{a}, \tilde{b}, \tilde{c} \in R \setminus 0$ with $a/b \sim \tilde{a}/\tilde{b}$ and $c/d \sim \tilde{c}/\tilde{b}$ is well-defined and gives the inverse elements.

Exercise 2.3.4. Let R be a ring and $\partial\colon R[x] \to R[x] : R \to 0, x \mapsto 1$ the standard derivative. Show that

a. if $p_1 \in R[x]$ is a polynomial, $p := (x-a)p_1 \in R[x]$ with $a \in R$, then ∂p has root a iff p_1 has root a;

b. conclude that for $p \in F[x]$ where F is a field, then the roots of $\gcd(p, \partial p)$ are exactly the multiple roots of p. (This will be helpful in the section about Discriminants of polynomials.)

Exercise 2.3.5. Let R be a domain and denote $F := K[R]$ the field of fractions of R.

a. Show that $K[R[x]] = F(x)$ where x is an indeterminate over R and $F(x)$ is the field of rational functions p/q for $p, q \in F[x]$ and $q \neq 0$.

b. Show that $K[R[x_1, x_2, \dots]] = F(x_1, x_2, \dots)$ where $F(x_1, x_2, \dots)$ is correspondingly the field of rational functions in several variables, i.e. p/q for $p, q \in F[x_1, x_2, \dots]$ and $q \neq 0$.

c. Show that $F(\!(x)\!) := K[R[[x]]] = F[[x], x^{-1}]$ where $F[[x], x^{-1}]$ are the formal Laurent series, i.e. the power series starting with a finite integer possibly negative exponent.

   Hint: Remember the geometric series, i.e. for $|q| < 1$, $\frac{1}{1-q} = 1 + q + q^2 + \dots$ and use this to invert a formal power series (in terms of power series with finite coefficients).

d. Show that the embedding $R[x] \to R[[x]]$ induces an embedding $F(x) \to F(\!(x)\!)$ that maps a rational function to a Laurent series. What element is $1/(1 + x + x^2) \in F(x)$ mapped to?

Exercise 2.3.6. Let $(R, +, \cdot, 1)$ be a ring. Show that for any unit polynomial $u \in R[x]^*$ (i.e. there is a $v \in R[x]$ such that $uv = 1 = vu$) its constant term is a unit in R and the highest non-zero coefficients are zero divisors (if they are not the constant term). What can you say about the coefficients in between?
Hint: You could start with quadratic (or one cubic) polynomials to see what happens.

Exercise 2.3.7. Given a field F of characteristic 0 (e.g. rational numbers). Show that the Taylor series

$$Tp := a_0 + \sum_{n \geq 1} \frac{a_n}{n!} x^n$$

with $a_n := \mathrm{ev}_0(\partial^n p)$ stops after finitely many steps for every polynomial $p \in R[x]$ where $\partial F = 0$ and $\partial x = 1$.

Exercise 2.3.8. Let $(R[x], \partial)$ be a differential ring and R be an integral domain. Show that ∂ extends uniquely to $K[R[x]] = F(x)$ with $F = K[R]$ and $F(x)$ as in Exercise 2.3.5-a. Express the constants $\mathrm{Const}(F(x))$ in terms of $\mathrm{Const}(R[x])$.
Hint: Show the quotient rule using the product/Leibniz rule.

Exercise 2.3.9. Let $p, d \in R[x]$ be polynomials over a domain R with $\deg d \geq 1$ and the leading coefficient of d in R^*. Show that there are unique polynomials $q_0, q_1, \dots \in R[x]$ with $\deg q_i < \deg p$ such that $p = q_0 + q_1 d + q_2 d^2 + \dots$.

Exercise 2.3.10. Let A be a noncommutative unital ring.

a. Find an example A, $a \in A$, and $p, q \in A[x]$ the naive polynomials such that $ev_a(pq) \neq \mathrm{ev}_a(p)\,\mathrm{ev}_a(q)$

b. Show that for "enhanced" polynomials $p, q \in T'(A)$ this does not happen, i.e. $\mathrm{ev}_a \colon T'(A) \to A : c_0 \otimes c_1 \otimes \dots \otimes c_n \mapsto c_0 a c_1 a \dots a c_n$ and linearly extended, then ev_a is an algebra homomorphism for every $a \in A$.

Exercise 2.3.11. Let $\mathfrak{m} \lhd R$ be a maximal ideal.

a. Show that $\mathfrak{m} + (x) \lhd R[x]$ is a maximal ideal in the polynomial ring over R.

b. Show that $\mathfrak{m} + (x_1, x_2, \dots, x_n) \lhd R[x_1, x_2, \dots, x_n]$ is a maximal ideal in the polynomial ring over R in several indeterminates.

c. Show also that $(x_1, x_2, \dots, x_n) = (x_1) + \dots + (x_n) \lhd R[x_1, x_2, \dots, x_n]$ the ideal generated by the n variables is not a principal ideal for $n \geq 2$.

d. Show that also $\mathfrak{m} + (x) \lhd R[[x]]$ is a maximal ideal.

Exercise 2.3.12. Let $R' := R[x_1, x_2, \dots, x_n]$ be a ring in several indeterminates. We call a polynomial $p \in R'$ homogeneous iff all its non-zero monomials have the same total degree. E.g. $x_1 x_2 + 2x_3^2$ is homogeneous of degree 2 while $x_1^2 - x_2^3$ is not homogeneous. Show that every polynomial $p \in R'$ can be written uniquely as a minimal sum of homogeneous polynomials.[4]

Exercise 2.3.13. Let R be a ring and consider $p, q \in R[[x]]$ two formal power series. Show that if q has no absolute term ($q_0 = 0$), then $p \circ q := \sum_{n \geq 0} p_n q^n$ is a formal power series where $p = p_0 + p_1 x + p_2 x^2 + \dots$ and $q = q_0 + q_1 x + q_2 x^2 + \dots$ respectively.
Hint: You have to show that the infinite sum in the definition gives only finitely many terms for every particular coefficient of $p \circ q$.

[4] Minimality w.r.t. the number of summands.

Exercise* 2.3.14. Develop a theory of formal power series in several variables over a (commutative) ring R with unit 1.

a. Define $R[\![x_1, x_2, \dots]\!]$ together with operations addition and multiplication such that there is an embedding $R[x_1, x_2, \dots] \to R[\![x_1, x_2, \dots]\!]$ that is a ring homomorphism.

b. Define substitution $\mathrm{subs}_{x\to t}\colon R[\![x]\!] \to R[\![x]\!]$ for $t \in R[\![x]\!]$ with vanishing absolute term, such that it extends $\mathrm{subs}_{x\to t}\colon R[x] \to R[x]$ for $t \in R[x]$ with vanishing absolute term and remains a ring homomorphism.

c. Find an expression for $K[R[\![x_1, x_2]\!]]$ in terms of $F := K[R]$, $F(\!(x_1)\!) = F[\![x_1]\!][x_1^{-1}]$, $F'(\!(x_2)\!)$ where F' may be any field.

2.4 Principal ideal domains (主理想环)

The integers have another more particular property, namely every ideal is already generated by one element. In order to see this, let $I \triangleleft \mathbb{Z}$ be any ideal and let $a \in I$ the smallest positive element. If no such element exists, then $I = \{0\} = (0)$. In the other case we claim $I = (a) := a\mathbb{Z}$. Let therefore $b \in I$ be any other element. It is clear from elementary number theory that the greatest common divisor $\gcd(a, b) = af_a + bf_b$ for some integers $f_a, f_b \in \mathbb{Z}$. Linear combination of a and b and is also an element in I. Making the gcd positive, we have either $0 < \gcd(a, b) < a$ or $a \leq \gcd(a, b)$. In the former case a is not the smallest positive element in I, if in the latter case $a < \gcd(a, b)$, then a cannot be an integer multiple of the gcd and thus the gcd not a divisor of a (each a contradiction). Therefore $a = \gcd(a, b)$ and thus b a multiple of a.

Considering on the other hand the polynomials in two variables $\mathbb{C}[x, y]$ with coefficients in $\mathbb{C}$ it is not hard to see that $I := (x, y) = (x) + (y)$ is an ideal that is not generated by only one element, because this element must divide x and y and be of total degree be at least 1, because $1 \notin I$.

Definition 2.4.1. *A domain R is called a* principal ideal domain (PID, 主理想域) *if every ideal $I \triangleleft R$ is generated by one element.*

Example 2.4.2. Given a field F, then the polynomials in one variable form a principal ideal domain. The ideal generated by k elements $(p_1, \dots, p_k)$ is the ideal generated by its greatest common divisor (see the Euclidean algorithm below).

Proposition 2.4.3. *Given a PID R, then every two elements a, b have a greatest common divisor d that is $Ra + Rb = Rd$.*

Proof. Note that for any divisor $d|a$ and $d|b$ we have $a, b \in (d)$. Conversely $Ra + Rb$ is an ideal in a PID. Therefore there is a $d \in R$ with $Ra + Rb = Rd$. Since $a, b \in Rd$ they are both multiples of d. On the other hand $d = fa + gb$ for some $f, g \in R$ and so every common divisor d' of a and b is also a divisor of d. □

How do we effectively compute the gcd? In the context of the following definition there is a very efficient algorithm.

Definition 2.4.4. *A Euclidean domain is a domain $(R, +, \cdot)$ together with a map* $\deg\colon R \setminus 0 \to \mathbb{N}$ *such that for every $a, b \in R$ with $b \neq 0$ there are $q, r \in R$ (quotient and remainder) with* $\deg r < \deg b$ *or $r = 0$ with*

$$a = qb + r$$

Example 2.4.5. 0. The integers are a Euclidean domain and the degree is the absolute value, i.e. $\deg z = |z|$.

1. The polynomials over a field are another example of Euclidean domains, because the degree defined as $\deg(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0) = n$ for $a_n \neq 0$ matches with polynomial division, i.e. the polynomial division stops if $\deg r < \deg b$.

   Note that in this case in addition $\deg(pq) = (\deg p) + (\deg q)$.

5

Proposition 2.4.6 (Euclid)**.** *Given a Euclidean domain together with two nonzero elements $a, b \in R$, then there is a common divisor for a, b such that* $\deg d$ *is maximal and every other common divisor divides d. Moreover the* gcd *can be written as a linear combination of a and b.*

Proposition 2.4.7. *Given a principal ideal domain, then the proper prime ideals are the ideals generated by irreducible elements, and these ideals are maximal.*

Proof.

Lemma 2.4.8. *Given a PID R and an irreducible element $p \in R$ that divides a product ab, $a, b \in R$, then either $p|a$ or $p|b$.*

Proof. Assume that p does not divide a. Then $\gcd(p, a) = 1 = fp + ga$. But then also $b = bfp + gab$ and each product is divisible by p. □

Therefore the irreducible elements generate prime ideals. Conversely every prime ideal is generated by one element $p \in R$. If p were not irreducible, we would

[5] Exercise: Argue that every element of degree 0 that is not 0 is a unit.

Input: two ring elements $a, b \in R$
Output: common divisor of highest degree of a and b together with the factors for the linear combination.
$f_a, f_b \leftarrow 0, 1$ – factors for b;
$g_a, g_b \leftarrow 1, 0$ – factors for a;
while $a \neq 0$ **do**
  Determine $q, r \in R$ such that $b = aq + r$ and $\deg r < \deg a$ or $r = 0$;
  $b \leftarrow a$;
  $a \leftarrow r$;
  $(f_a, g_a) \leftarrow (g_a, f_a - qg_a)$;
  $(f_b, g_b) \leftarrow (g_b, f_b - qg_b)$;
end
return (b, f_a, f_b). □

Algorithm 1: Euclidean algorithm

have $p = ab$ with both $a, b \in R$ no units and thus $ab \in Rp$, but $a + Rp \neq R$ and $b + Rp \neq R$, i.e. the quotient ring is not a domain.

The prime ideals Rp are also maximal, because every ideal properly containing Rp is of the form Rd where $d|p$. But p is irreducible, so d is a unit and thus $Rd = R$. □

In preparation of the next chapter, note the following property of PIDs.

Proposition 2.4.9 (ACC)**.** *Given a PID R, then every ascending chain of ideals $\{0\} \lhd I_1 \lhd I_2 \lhd I_3 \ldots$ is finite, i.e. there is an $n \in \mathbb{N}$ such that $I_{n+1} = I_{n+2} = I_{n+3} = \ldots$.*

Proof. Consider I_∞ the union of all the ideals. It is clearly another ideal. Since R is a PID it is generated by one element $a \in R$. Since a is in the union of the ideals, there is an n with $a \in I_n$. Then clearly $I_\infty \lhd I_n$, but then $(a) = I_\infty \lhd I_n \lhd I_{n+1} \lhd I_{n+2} \lhd \ldots \lhd I_\infty$ and therefore all the later ideals are the same. □

2.4.1 Rational functions (有理函数)

An important application of PIDs is the decomposition of an arbitrary rational function $f \in F(x)$ into a sum of partial fractions $f = p + p_{1,1}/q_1 + p_{1,2}/q_1^2 + \cdots + p_{n,k_n}/q_n^{k_n}$ where the q_i are irreducible and $\deg p_{i,j} < \deg q_i$. If you are working over an algebraically closed field F, then all you need to do is to introduce elementary transcendental functions $\log(x - r)$ integrating the $1/(x - r)$ in order to be able to integrate arbitrary rational functions. If F is not algebraically closed (such as $\mathbb{R}$), you may have to introduce a few more transcendental functions, but the principle is the same.

Theorem 2.4.10. *Every rational function $f \in F(x)$ over a field can be written uniquely as a sum of a polynomial $p \in F[x]$ and partial fractions $p_{i,j}/q_i^j$ with $\deg p_{i,j} < \deg q_i$ and every q_i divides the denominator of f.*

The proof covers the main part of this subsection and consists of a series of lemmas.

Lemma 2.4.11. *For every rational function $f \in F(x)$ there is a unique reduced form $f = p/q$ with $p, q \in F[x]$, p and q have no non-constant common factor, and q monic.* □

Lemma 2.4.12. *Every rational function $f \in F(x)$ can be written uniquely as $f = p + p_1/q$ with a polynomial $p \in F[x]$ and a proper fraction p_1/q where $p_1, q \in F[x]$ have $\gcd(p_1, q) = 1 \in F$, q is monic, and $\deg p_1 < \deg q$.* □

We call a rational function $f \in F(x)$ polynomial free if $f = p/q$ with $p, q \in F[x]$ has $\deg p < \deg q$. Note that in this case the polynomial part in the above decomposition vanishes (i.e. is 0).

Lemma 2.4.13. *Given a polynomial free rational function $f = p/q \in F(x)$ with $p, q \in F[x]$ relatively prime and q monic. Then f can uniquely be written in the form*

$$f = \frac{p_1}{q_1^{k_1}} + \cdots + \frac{p_n}{q_n^{k_n}}$$

where $q_1, \ldots, q_n$ divide q, are irreducible, their degrees add up to the degree of q: $k_1 \deg q_1 + \cdots + k_n \deg q_n = \deg q$, and $\deg p_i < k_i \deg q_i$. □

Lemma 2.4.14. *Given a polynomial free rational function $f = p/q^k \in F(x)$ where $\deg q \geq 1$, then there exist unique polynomials $p_1, \ldots, p_k \in F[x]$ with $\deg p_j < \deg q$ and*

$$\frac{p}{q^k} = \frac{p_1}{q} + \frac{p_2}{q^2} + \cdots + \frac{p_k}{q^k} \quad \square$$

2.4.99 Exercises

Exercise 2.4.1. Compute the gcd and lcm of x^2+x-1, x^3+x-1, and x^4+x^2-1 over $\mathbb{Q}$.

Exercise 2.4.2. Show that no polynomial ring in more than one indeterminate is a PID.

Exercise 2.4.3. The Gauss integers are the ring $\mathbb{Z}[\mathrm{i}] \subset \mathbb{C}$ with the usual imaginary unit $\mathrm{i}^2 = -1$. Show that $\mathbb{Z}[\mathrm{i}]$ is a PID.
Hint: You could show that for every $a, b \in \mathbb{Z}[\mathrm{i}]$ and $b \neq 0$ there are $q, r \in \mathbb{Z}[\mathrm{i}]$ with $a = qb + r$ and $|r| < |b|$.

Exercise 2.4.4. Show that for every family $(a_i)_{i\in I}$ of elements $a_i \in R$ of a PID the greatest common divisor can be written as finite linear combination $\gcd = \sum_{j=1}^{n} c_j a_{i_j}$ for some $i_j \in I$, $n \in \mathbb{N}$ and $c_j \in R$.

Exercise 2.4.5. Consider polynomials over the rational numbers $\mathbb{Q}$.

a. Write $p = x^5 + x^3 - x^2 - 1 \in \mathbb{Q}[x]$ as a product of irreducible polynomials.

b. Do the same for $p = x^4 + 1 \in \mathbb{Q}[x]$.

Exercise 2.4.6. Write down all irreducible polynomials in $\mathbb{F}_2[x]$ of degree 5.

Exercise 2.4.7. Prove the lemmas 2.4.11–2.4.14.

Exercise 2.4.8. Write in partial fractions

$$\frac{x^5+1}{x^4+x^2} \in \mathbb{F}_2(x), \tag{a}$$

$$\frac{x^5+1}{x^4+x^2} \in \mathbb{F}_3(x), \tag{b}$$

$$\frac{1}{x^5+x^3+x} \in \mathbb{F}_2(x). \tag{c}$$

2.5 Unique factorization domains (唯一分解整环)

The PIDs have also another interesting property.

Definition 2.5.1. *Given a domain R. It is called a* unique factorization domain *(UFD, 唯一分解整环) if every nonzero non-unit element has a (finite) factorization into irreducible elements and for every two factorizations of a nonzero non-unit element $a \in R \setminus (\{0\} \cup R^*)$ into irreducible elements $a = p_1 \dots p_m = q_1 \dots q_n$ there is a bijection σ between the p_is and q_js together with units $u_i \in R^*$ such that $p_i = u_i q_{\sigma i}$.*

This means that factorizations into irreducibles are unique up to rearrangements and multiplications of the factors with units.

Proposition 2.5.2. *Given a PID R, then it is a unique factorization domain.*

Proof. First we need to argue that every nonzero non-unit element $a \in R$ has a factorization into irreducible elements. We start therefore with its factorizations. If all factorizations into two factors contain at least one unit, we are done as a is irreducible. If a has nontrivial factorizations $a = bc$ this means in particular that $(a) = (b)(c)$ and thus $(a) \lhd (b), (c)$. We continue by factorizing b and c until we

end up with an irreducible element. Due to Proposition 2.4.9 all chains of factors of ideals generated this way must be finite. Therefore a breaks into finitely many irreducible factors.

In order to prove uniqueness of the factorization let now $a = p_1 \dots p_m = q_1 \dots q_n$ where p_i and q_j are irreducible. Start with $i = 1$ and observe that either $p_1|q_1$ or $p_1|q_2 \dots q_n$ by Lemma 2.4.8. In the first case $q_1 = u_1 p_1$, choose $\sigma(1) := 1$ and note that $u_1 \in R^*$, because q_1 is irreducible. In the second case go over the other factors of $q_2 \dots q_m$ in order to find $\sigma(1)$ and u_1. In particular the second product cannot be empty ($n > 0$), because p_1 is not a unit. Conversely after we have assigned $\sigma(m)$, the remaining factors q_j must multiply to 1 and are therefore units. Thus $m = n$. □

Example 2.5.3.

1. Therefore the integers as well as the polynomials with coefficients in a field form a UFD.

2. But also Gauss integers $\mathbb{Z}[\mathsf{i}] \subset \mathbb{C}$ with the usual convention $\mathsf{i}^2 = -1$ form a UFD, because they are a Euclidean domain with $\phi\colon \mathbb{Z}[\mathsf{i}] \to \mathbb{N} : a+b\mathsf{i} \mapsto a^2+b^2$ where $|z|^2 = z\bar{z}$ and $\overline{a+b\mathsf{i}} = a - b\mathsf{i}$ is the complex conjugate.

3. Correspondingly $\mathbb{Z}[\sqrt{-2}]$ is a UFD.

4. $\mathbb{Z}[\sqrt{-5}]$ is not a UFD. For example $9 = 3 \cdot 3 = (2+\sqrt{-5})(2-\sqrt{-5})$, but as we can see from the norm $N(a+b\sqrt{-5}) := a^2+5b^2$ with $N(wz) = N(w)N(z)$ and $N(3) = 9 = N(2 \pm \sqrt{-5})$ all three factors are irreducible without any unit $u \in \mathbb{Z}[\sqrt{-5}]$ such that $3u = 2 \pm \sqrt{-5}$.

2.5.1 Irreducible Polynomials

The interesting but difficult question whether a polynomial over a field is irreducible was first answered with the following proposition due to F.G.M. Eisenstein.[6]

Proposition 2.5.4 (Eisenstein's criterion)**.** *Given a field $F = K[R]$ that is the field of fractions of a unique factorization domain R. Given further a polynomial $Q = a_0 + \dots + a_n x^n \in R[x]$ and an irreducible $p \in R$ such that*

1. *a_n is not divisible by p,*

2. *$a_0, \dots, a_{n-1}$ are divisible by p,*

3. *a_0 is not divisible by p^2,*

[6] *4/1823 †10/1852

then Q is irreducible in $F[x]$.

Proof. Assume in the sense of an indirect proof that Q fulfills the conditions of the criterion but also factors as $Q = fg$ with $f, g \in F[x]$ both of degree at least 1. By extracting the least common denominator $d \in R$ from the coefficients of f and g in relatively prime quotients, we can write $Qd = FG$ with $F, G \in R[x]$. Consider now the localizations of F and G modulo p, i.e. the polynomial $Q_p := [a_0] + [a_1]x + \cdots + [a_n]x^n$ with $[a_k] \in R/(p)$ and analogously for F_p and G_p. Note that $[a_n d] \neq [0]$, because a_n was not divisible by p, but all its other coefficients are. Starting with $[a_0 d] = [f_0][g_0]$ we know that $[f_0] = [0]$ or $[g_0] = [0]$. By induction over the coefficients of Q_p all the lower coefficients of F_p and G_p have to vanish. But then $a_0 = f_0 g_0$ is divisible by p^2 in contradiction to our assumption. □

Example 2.5.5 (Primitive cyclotomic polynomials)**.** Consider the polynomial $f_0 := x^p - 1 \in \mathbb{Z}[x]$. This can be factored as $f_0 = (x-1)(x^{p-1} + x^{p-2} + \cdots + 1)$. Denote the second factor as $g \in \mathbb{Z}[x]$ and let $p \in \mathbb{P}$ be a prime. Note that we cannot yet apply Eisenstein's criterion to g with our p. Instead consider $f := g(x+1) = [(x+1)^p - 1]/x = x^{p-1} + \binom{p}{p-1}x^{p-2} + \cdots + \binom{p}{2}x + \binom{p}{1}$. Observe that now p does not divide $a_{p-1} = 1$, but does divide $a_{p-2} = \frac{p!}{(p-1)!\,1!}$,..., $a_0 = p$. Moreover p^2 does not divide $a_0 = p$. Therefore f is irreducible in $\mathbb{Q}[x]$ and thus also g. The polynomials g are called primitive cyclotomic polynomials. They play an important role in Galois theory.

Note that $x^n - 1 = (x-1)\prod_{p^k|n} \frac{x^{p^k}-1}{x^{p^{k-1}}-1}$ where the product runs over all prime power divisors of n. The shown irreducibility shows that this is the factorization into irreducibles.

The problem is that the criterion cannot be applied to every irreducible polynomial. Using more localizations one can derive better irreducibility tests.

The following result is already due to Euclid:

Proposition 2.5.6 (Euclid)**.** *Given a UFD R, then $R[x]$ contains infinitely many different irreducible polynomials.*

Remark 2.5.7. All the polynomials of degree 1 are irreducible. If R itself has infinitely many elements, then the $x - r \in R[x]$ for $r \in R$ are all different.

Proof. Assume for the sake of an indirect proof that $p_1, \ldots, p_N$ are all different irreducible polynomials. Consider now their product $P := 1 + \prod_{k=1}^{n} p_k$. Since the list contains at least the linear monic polynomials the product has degree at least 2. Thus it is not a unit and not 0. But none of the p_k divides P, because it has remainder 1. On the other hand P does factor into a product of irreducible polynomials. Therefore there must be at least one other polynomial p_{n+1} that divides P. This is a contradiction to an assumed finite list. □

Theorem 2.5.8. *Given a UFD R, then the polynomials over R (in one more indeterminate) form a UFD.*

The first proof of the theorem was discovered by Gauss in the case of $\mathbb{Z}[x]$ and uses the following notions:

Definition 2.5.9. *Given a polynomial $p \in R[x]$ where R is a UFD. We say that p is* primitive *if there is no irreducible $q \in R$ such that q divides all coefficients of p.*

The following lemma is immediate and its result will help in the proof of the Theorem.

Lemma 2.5.10. *Every non-zero polynomial $p \in F[x]$ where $F = K[R]$ is the field of fractions of a UFD R can be written in the form $p = up^*$ where $u \in F$ and $p^* \in R[x]$ primitive.* □

We denote $u \in F[x]$ the *content* of p. The content of a primitive polynomial is thus 1 (or any other unit $u \in R^*$).

Lemma 2.5.11 (Gauss)**.** *The product of two primitive polynomials is primitive.*

Proof. Let $p, q \in R[x]$ be primitive polynomials over the UFD R with coefficients $p = p_0 + p_1x + \cdots + p_mx^m$, $q = q_0 + q_1x + \cdots + q_nx^n$. And let $r \in R$ be any irreducible element. Since p and q are primitive, not all of the p_i, q_i are divisible by r. Let k be the smallest i such that r does not divide p_i and l the smallest i such that r does not divide q_i. But then r divides all p_i with $i < k$ and all q_j with $j < l$. Thus in particular for $pq = c_0 + c_1x + \cdots + c_{m+n}x^{m+n}$, p does not divide $c_{k+l} = \sum_{i<k, j<l} p_iq_j + p_kq_l$. □

Proposition 2.5.12. *A polynomial $p \in F[x]$ over the field of fractions of a UFD R, $F = K[R]$, then p is irreducible iff $p^* \in R[x]$ is.*

Proof. We may assume that $\deg p \geq 1$. If p is not irreducible over $F[x]$, then there are $f, g \in F[x]$ with $p = fg$. But then also $p^* = (fg)^* = f^*g^*$ by the last lemma and thus p^* factors over $R[x]$.

If on the other hand $p^* = fg$ for some $f, g \in R[x]$. Since p^* is primitive, so must be f and g. In particular each of them has degree at least 1. But then also $p = up^* = ufg$ is not irreducible in $F[x]$. □

Remark 2.5.13. To say it more explicitly, the irreducible elements of $R[x]$ are the irreducible elements of R together with the $p^* \in R[x]$ where $p \in F[x]$ is an irreducible polynomial. In particular all the p^* are primitive.

Proof of the theorem. We work by induction over the number of indeterminates. Let us thus restrict to the case of one indeterminate. In order to show that $R[x]$ is a UFD, we consider $F := K[R]$, because R is in particular a domain, and embed the polynomials $R[x] \subset F[x]$. In the latter case we have already shown (see Proposition 2.5.2) that these form a UFD. Assume thus that $p \in R[x] \subset F[x]$ factors as $p = p_1 \dots p_n$ with $p_i \in F[x]$. By extracting the *content*, we can write $p = uP_1 \dots P_n$ where $P_i \in R[x]$ and $p_i = u_i P_i$, $u = u_1 \dots u_n \in F$. If $u_i \in R^*$, then also $p_i \in R[x]$ and thus we can keep this factor. If there are some $p_i \notin R[x]$ it means that also the corresponding $u_i \notin R$. We are thus left with factoring the polynomials p for which all $p_i \notin R[x]$, but then the polynomial is irreducible over R, while we claim that it splits over $F[x]$. This is a contradiction to Lemma 2.5.12.

It remains to show uniqueness of the factorization. Let thus $p = p_1 \cdots p_m = q_1 \cdots q_n$ with $p_i, q_i \in R[x]$. From the uniqueness of factorization over $F(x)$ we know that there is a bijection between the polynomials that are not constant, i.e. $p_i = u_i q_{\sigma i}$ for $u_i \in F^*$ and $1 \leq i \leq m' \leq m$. If the p_i are irreducible over R, then they must be either constant or of degree at least 1 and primitive. But then the $u_i \in R^*$ and the remaining $p_{m'+1} \cdots p_m$ as well as the remaining $q_{m'+1} \cdots q_n$ must multiply to an element of F^*. Since all p_i are primitive and of degree at least 1 and p^* is primitive as well, we have $p^* = up_1 \cdots p_{m'} = u' q_{\sigma 1} \cdots q_{\sigma m'}$ as well as the content of p equal some element in R, $u^{-1} p_{m'+1} \cdots p_m = u'^{-1} q_{\sigma(m'+1)} \cdots q_{\sigma n}$. But since R is a UFD, we know that we can modify the bijection σ such that $p_i = u_i q_{\sigma i}$ also for $m' + 1 \leq i \leq m$, $u_i \in R^*$ and in particular $n = m$. This completes the proof. □

Example 2.5.14.

1. We already know that the integers $\mathbb{Z}$ form a UFD, because they are a PID. Therefore also $\mathbb{Z}[x]$, the polynomials with integer coefficients, form a UFD. Note that they are no longer a PID.

2. In the same way polynomials in arbitrary many indeterminates over a UFD (e.g. a field) form a UFD, because every (finite) polynomial lies in an extension with finitely many indeterminates which form a tower of UFDs.

2.5.99 Exercises

Exercise 2.5.1. Show that every family of elements $a_i \in R$ of a UFD has a

a. greatest common divisor;

b. least common multiple, if the family is finite;

c. show that there is an infinite family of elements that do not have a finite non-zero least common multiple.

Hint: You cannot assume that an UFD is a PID, but nevertheless the gcd is determined uniquely by the common irreducible divisors modulo equivalence, while the lcm is determined by the union of all irreducible divisors modulo equivalence.

Exercise 2.5.2. Find a UFD R together with two elements $a, b \in R$ such that their greatest common divisor cannot be written as a linear combination $\gcd(a, b) = fa + gb$ for any $f, g \in R$.

Exercise 2.5.3. Prove the following: Assume R is a UFD, $\mathfrak{p} \lhd R$ a prime ideal and $\pi\colon R \to R/\mathfrak{p}$ the quotient map. If $f \in R[x]$ monic and $f_{\mathfrak{p}} \in (R/\mathfrak{p})[x]$ irreducible, then f is irreducible over R.

Exercise 2.5.4. Apply the previous result to show that the following polynomials are irreducible in $\mathbb{Q}[x]$:

a. $x^3 - 10$,

b. $x^3 + 3x^2 - 6x + 3$,

c. $x^3 + 3x^2 - 6x + 9$,

d. $x^3 - 3x + 4$.

2.8 Localizations (环的局部化)

Remember the trick in the proof of Eisenstein's criterion (Proposition 2.5.4). Given a polynomial $Q \in R[x]$ over an integral domain R together with a prime ideal $\mathfrak{p} \lhd R$, then Q reducible in $R[x]$ implies that $Q_{\mathfrak{p}}$ is reducible in $(R/\mathfrak{p})[x]$, because the projection $\pi\colon R[x] \to (R/\mathfrak{p})[x] : a \mapsto a + \mathfrak{p}, x \mapsto x$ can be extended as a ring homomorphism. Therefore, given a prime ideal $\mathfrak{p} \lhd R$ such that $Q_{\mathfrak{p}}$ is irreducible, then also Q must be irreducible.

Example 2.8.1.

You will see in the homework what can be done if there is no prime ideal $\mathfrak{p} \lhd R$ such that $Q_{\mathfrak{p}}$ is irreducible, but you have the strong feeling that Q should be irreducible.

A related construction is that of a field of fractions. As we saw in the subsection about irreducible polynomials, the irreducible elements over the field of fractions are closely related to the irreducible elements in the polynomial extension over the base ring.

The biggest problem that prevents us from constructing the quotient field of an arbitrary (commutative) ring R are the zero divisors. Namely if we write a/b

for some $a, b \in R$ we do not only have to exclude $b = 0$, but also every possible $d_{1/2}$ such that $d_1 d_2 = 0$, i.e. the zero divisors, because they would eventually lead to $(1/d_1)(1/d_2) = 1/0$.

The abstraction gives the following concept.

Definition 2.8.2. *Given a ring $(R, +, \cdot)$, then a multiplicative set is a nonempty subset $S \subset R$ that does not contain 0, contains all units $R^* \subset S$, and is closed under multiplication, i.e. $S \cdot S \subset S$.*

Example 2.8.3.
1. The complements $S = R \setminus \mathfrak{p}$ of prime ideals $\mathfrak{p} \lhd R$ are multiplicative sets, because for a prime ideal we have that $ab \in \mathfrak{p}$ implies $a \in \mathfrak{p}$ or $b \in \mathfrak{p}$ and in particular $0 \in \mathfrak{p}$ thus $0 \notin S$.

2. For domains (such as the integers $\mathbb{Z}$) (0) is a prime ideal and we can thus choose $S = R \setminus \{0\}$.

The localization is now defined as follows.

Definition 2.8.4. *Let R be a ring and $S \subset R$ be a multiplicative subset. The localization of R at S, denoted as $S^{-1}R$ is the set $R \times S/\sim$ with elements denoted as a/s where $a \in R$ and $s \in S$ under the equivalence relation $r/s \sim r'/s'$ iff there is a $t \in S$ such that*

$$t(rs' - r's) = 0.$$

Addition is done via expansion to a common denominator, i.e. $a/s + b/t = (at + bs)/(st)$. Multiplication is done componentwise, i.e. $(a/s)(b/t) = (ab)/(st)$.

Proposition 2.8.5. *The relation $\sim$ is an equivalence relation. $a/s \sim (at)/(st)$ and addition and multiplication are representation independent and thus $S^{-1}R$ a ring together with a ring homomorphism $i\colon R \to S^{-1}R : a \mapsto a/1$ that maps all elements of $S \subset R$ to units.*

It is obvious that this generalizes the definition of the field of fractions in Proposition 2.3.3. The difference to the quotient field of a domain is that R can have zero divisors.

Proof. Reflexivity $a/s \sim a/s$ is clear with any $t \in S$. Suppose $a/s \sim b/s'$ with $t \in S$, then $t(bs - as') = -t(as' - bs) = 0$, and for transitivity assume that $a/s \sim a'/s'$ via $t \in S$ and $a'/s' \sim a''/s''$ via $t' \in S$, then $tt's'(as'' - a''s) = t's''(t(as' - a's)) + ts(t'(a's'' - a''s')) = 0$. The second statement is obvious.

In order to see that addition is representation independent, let $a/s \sim a'/s'$ via t and $b/u \sim b'/u'$ via t'. Then $a/s + b/u = (au + bs)/(su)$ and $a'/s' + b'/u' = (a'u' + b's')/(su)$. The right hand expressions are equivalent via tt', because $tt'[(au + bs)s'u' - (a'u' + b's')su] = tt'[(as' - a's)uu' + (bu' - b'u)ss'] = 0$. An

analogous but simpler computation leads to representative independence of the multiplication.

It is obvious that i is a ring homomorphism. The inverse elements of s are $1/s$. This completes the proof. □

Proposition 2.8.6. *The above construction fulfills the universality condition that every ring homomorphism $\phi\colon R \to R'$ that sends S to R'^* extends uniquely to $S^{-1}R$.*

Proof. Remember that the elements of $S^{-1}R$ are generated by pairs a/s with $a \in R$ and $s \in S$. Since $\phi(s) \in R'^*$ there are inverse elements $\phi(s)^{-1} \in R'$. We thus send $a/s \mapsto \phi(a)\phi(s)^{-1}$. It remains to check that this is representation independent. Let thus $a/s \sim a'/s'$ via $t \in S$. We thus have $t(as'-a's) = 0 \in R$ but since ϕ is a ring homomorphism, we obtain $\phi(t)(\phi(a)\phi(s') - \phi(a')\phi(s)) = 0 \in R'$. Since $\phi(t)$, $\phi(s)$, and $\phi(s')$ are invertible, we obtain $\phi(a)\phi(s)^{-1} = \phi(a')\phi(s')^{-1}$. The homomorphism property of the extension of ϕ is now obvious. This completes the proof. □

2.8.99 Exercises

Exercise 2.8.1. Given the ring of integers $R = \mathbb{Z}$ and the multiplicative set $S = R \setminus (2)$. Determine the localization $S^{-1}R$ and show that it is embedded into $K[R] = \mathbb{Q}$. What is its image?

Exercise 2.8.2. Show that the polynomial $Q = x^4 + 4x^3 + 5x^2 + 1 \in \mathbb{Z}[x]$ is irreducible in $\mathbb{Q}[x]$. You may proceed as follows:

0. Note that you cannot apply Eisenstein's criterion directly;

a. localize the polynomial w.r.t. $p = 2$ and note that it factors. This says, that one localization is not sufficient;

b. localize the polynomial w.r.t. $p = 3$ and note that it also factors. Thus this localization is not sufficient either;

c. assume that Q factors in $\mathbb{Q}[x]$ as $Q = fg$ and compare their localizations for $p = 2, 3$ with the factors you obtained earlier. Conclude that either f or g must have degree 4 and thus Q is irreducible;

Chapter 3

Field extensions (域扩张) and Galois theory (伽罗瓦理论, 5 weeks)

3.1 Algebraic and transcendental extensions (代数与超越扩张)

Remark 3.1.1. Note that a field F has only two ideals, (0) and F. Therefore every ring-homomorphism from a field is an embedding. So the main part in the study of fields is the study of field extensions.

Definition 3.1.2. *Let $F \subset E$ be a field extension. An element $\alpha \in E$ is called* algebraic over F (代数的) *if there is a nontrivial polynomial $p \in F[x]$ such that $p(\alpha) = 0$.*

An element that is not algebraic is called transcendental (超越).

A field extension $F \subset E$ is called algebraic if every element in E is algebraic over F.

A field extension that is not algebraic is called transcendental.

Example 3.1.3. 1. Consider the number $\sqrt{2} \in \mathbb{R}$. It is algebraic over $\mathbb{Q}$, because it is a root of $x^2 - 2 \in \mathbb{Q}[x]$. We can form the field extension $\mathbb{Q}(\sqrt{2})$ as the subalgebra in $\mathbb{R}$ generated by $\mathbb{Q}$ and $\sqrt{2}$. Since $\dim_{\mathbb{Q}} \mathbb{Q}(\sqrt{2}) = \deg(x^2-2) = 2$ we know that the increasing powers of any element $\alpha \in \mathbb{Q}(\sqrt{2})$ are linear dependent over $\mathbb{Q}$. Therefore $\mathbb{Q}(\sqrt{2})$ is an algebraic extension.

2. Consider $\mathbb{Q}(x)$ the field of rational functions over $\mathbb{Q}$ in one variable. Since x is a free variable, there is no nontrivial polynomial $p \in \mathbb{Q}[x]$ such that $p(x) \equiv 0$.

Therefore $\mathbb{Q}(x)$ is a transcendental extension of $\mathbb{Q}$ and x a transcendental element.

The same is true for $\mathbb{Q}(\mathrm{e})$ once you have proven that the Euler number e is transcendental.

3. Remember the characteristic of a ring. It is defined via the ring homomorphism $\phi_0 \colon \mathbb{Z} \to F : n \mapsto n \cdot 1$ where in the sense of this chapter we assume F to be a field. We already know that $\ker \phi_0 = (n) \lhd \mathbb{Z}$ a principal ideal in $\mathbb{Z}$. But moreover the image $\mathbb{Z}/(n) \cong \operatorname{im} \phi_0 \subset F$ is a subring in the domain F and thus also a domain. Therefore (n) must be a prime ideal, i.e. the characteristic of any field is either a prime number or 0. If $Z/(p) \subset F$, then this is the *prime field (基本域)* of F. If the characteristic is 0, then beside $\mathbb{Z} \subset F$ we also have its field of fractions lying in F, because F is a field, i.e. $\mathbb{Q} \subset F$. Thus in the case of characteristic 0, the field of rational numbers is the prime field. In what follows we often need a particular field extension $F \subset E$ and intermediate fields $F \subset K \subset E$. As soon as E is fixed, you can always choose the prime field as F, because it is automatically contained in every subfield $K \subset E$.

4. Note that in any field extension $F \subset E$, E is a vector space over F. If it is a finite dimensional vector space, we denote $[E : F] := \dim_F E$ the *algebraic degree (代数度数)* of E over F. Analogous to the first example every element in E must be algebraic over F. This justifies the name. Conversely if E/F contains transcendental elements, then its dimension must be infinite.

Proposition 3.1.4 (Tower property)**.** *Given algebraic extensions $F \subset K \subset E$, then $[E : F] = [E : K]\,[K : F]$ and in particular $[E : F]$ is infinite iff either $[E : K]$ or $[K : F]$ is infinite.*

Proof. Note that a K-base $\{e_i\}$ of E together with an F-base $\{k_j\}$ of K gives an F-base of E as follows $\{e_i k_j\}$. Thus the dimension formula follows. □

Definition 3.1.5. *Given a field F, then any $f \in F$ for which there is an $n \in \mathbb{N}_+$ with $f^n = 1$ is called a root of unity.*

Example 3.1.6. In the complex numbers the n-th roots of unity are $\mathrm{e}^{2\pi \mathrm{i} k/n}$ for $k = 0, \ldots, n-1$. They form a cyclic group with generators any primitive n-th root, e.g. $\omega_n := \mathrm{e}^{2\pi \mathrm{i}/n}$ or more generally any $\mathrm{e}^{2\pi \mathrm{i} k/n}$ as long as k and n are relatively prime.

The latter property is shared among all fields, i.e.

Proposition 3.1.7. *Every finite multiplicative subgroup of a field is cyclic and thus consists of roots of unity.*

Proof. Let F be the field in consideration and $G \subset F^*$ be a finite group of order $(G:1) = p_1^{k_1} \cdots p_n^{k_n}$. Since G is abelian, its structure is $G = G_1 \times \cdots \times G_n$ with G_k abelian and of order $p_k^{n_k}$ for distinct primes $p_k \in \mathbb{P}$ and positive integer exponents $n_k \in \mathbb{N}_+$. It remains to show that the G_k are all cyclic. Consider conversely the set of all $\{r \in G : r^{p^j} = 1 \exists j \in \mathbb{N}_+\}$. It consists exactly of G_k iff $p = p_k$. Since G and thus G_k is finite, there is an element $r \in G_k$ of maximum p-power say $K \in \mathbb{N}$. Clearly $\langle r \rangle \subset G_k$ is a cyclic subgroup with $p^K = \operatorname{ord} r \leq (G_k : 1)$. On the other hand the polynomial $x^{p^K} - 1 \in F[x]$ has at most p^K roots in F. But all the elements in G_k are roots of this polynomial and thus roots of unity. Therefore $(G_k : 1) \leq p^K$ and thus $\langle r \rangle = G_k$ and in particular cyclic. □

Corresponding to rings we have the following construction:

Lemma 3.1.8. *Given a family of subfields $\{F_\alpha : \alpha \in A\}$ of a field F, then their intersection $\bigcap_{\alpha \in A} F_\alpha \subset F$ is a sub-field of F.*

The union of a non-empty chain of sub-fields $F = F_0 \subset F_1 \subset F_2 \subset \cdots \subset E$ is a sub-field of E.

The proof is left as an exercise.

Definition 3.1.9. *Given a field extension $F \subset E$. The* sub-field generated by a subset *$S \subset E$ over F is the smallest field $\bigcap_{S, F \subset K \subset E} K \subset E$ that contains F and all elements of S. It is denoted as $F(S)$.*

Proposition 3.1.10. *Given a field extension $F \subset E$ and a subset $S \subset E$, then the ring $F[S]$ is the set of all finite F-linear combinations of products of elements in S.*

The sub-field $F(S)$ generated by F and S is the field of fractions of the domain $F[S]$, i.e. it consists of all the elements $a/b \in E$ with $a, b \in F[S]$ and $b \neq 0$.

Also this proof is left as an exercise.

The last proposition has some immediate consequences:

Corollary 3.1.11. *Given a field extension $F \subset E$, a subset $S \subset E$, and some elements $r, \alpha_1, \ldots, \alpha_n \in E$, then*

1. *$r \in F[\alpha_1, \ldots, \alpha_n]$ iff there is some polynomial in n indeterminates $p \in F[x_1, \ldots, x_n]$ such that $r = p(\alpha_1, \ldots, \alpha_n)$;*
2. *$r \in F(\alpha_1, \ldots, \alpha_n)$ iff there is some rational function in n indeterminates $f \in F(x_1, \ldots, x_n)$ such that $r = f(\alpha_1, \ldots, \alpha_n)$;*
3. *$r \in F[S]$ iff there are some $\alpha_1, \ldots, \alpha_n \in S$ such that $r \in F[\alpha_1, \ldots, \alpha_n]$;*
4. *$r \in F(S)$ iff there are some $\alpha_1, \ldots, \alpha_n \in S$ such that $r \in F(\alpha_1, \ldots, \alpha_n)$.*

Also the proof of this corollary is left as an exercise.

Remark 3.1.12. Given a transcendental field extension E/F, then we can analogous to vector spaces introduce a *transcendence degree (超越次数)* by counting the minimal number of (transcendental) elements required to generate E over F. Analogously to vector spaces this number is independent of the particular transcendence base.

Finally the products of subfields are the following:

Definition 3.1.13. *Given a family of sub-fields $F \subset K_\alpha \subset E$, then their* composite $\prod_{\alpha\in A} K_\alpha$ *is the subfield of E generated over F by $\bigcup_{\alpha\in A} K_\alpha$.*

Note that this notion is symmetric in the fields K_α. In particular if $F \subset K_{1/2} \subset E$ are two intermediate fields, then $K_1(K_2) = F(K_1, K_2) = K_2(K_1) \subset E$ and correspondingly for more factors.

3.1.99 Exercises

Exercise 3.1.1. a. Give a short proof to show that there is no field of order 6.

b. What can you say about fields of order 10, 12, 14?

c. Given the example $\mathbb{Q}(\sqrt{2}) \cong \mathbb{Q}(x)/(x^2-2)$ of degree 2 over $\mathbb{Q}$, what would you need to construct a field with 4, 8, 9, 16 elements?

Exercise 3.1.2. Show Lemma 3.1.8, i.e. given a field extension $F \subset E$ and

a. a family of sub-fields $F_\alpha \subset E$, then their intersection $\bigcap_{\alpha\in A} F_\alpha \subset E$ is a sub-field;

b. a directed chain of sub-fields $F = F_0 \subset F_1 \subset F_2 \subset \cdots \subset E$, then their union $\bigcup_{i=0}^{\infty} F_i \subset E$ is a sub-field.

Exercise 3.1.3. Show Proposition 3.1.10, i.e. given a field extension $F \subset E$ together with a subset $S \subset E$, then

a. the ring $F[S]$ consists of all finite F-linear combinations of products of elements of S;

b. the field $F(S)$ is the field of fractions of the domain $F[S] \subset E$.

Exercise 3.1.4. Assuming the previous exercise, show Corollary 3.1.11.

3.2 The algebraic closure (代数闭包)

Definition 3.2.1. *Given an algebraic element $\alpha \in E \supset F$ of a field extension $F \subset E$. A minimal polynomial for α is a non-constant polynomial $p \in F[x]$ such that $p(\alpha) = 0$ and $\deg p$ is minimal.*

Example 3.2.2. Consider $\sqrt{2} \in \mathbb{C} \supset \mathbb{Q}$. It is algebraic, because it is a root of $p = x^2 - 2 \in \mathbb{Q}[x]$. Since $\sqrt{2} \notin \mathbb{Q}$, 2 is the smallest degree of a non-trivial polynomial with root $\sqrt{2}$.

Proposition 3.2.3. *Given an algebraic element $\alpha \in E \supset F$ in a field extension $F \subset E$. Then the minimal polynomial $p \in F[x]$ of α is irreducible over F and every polynomial $g \in F[x]$ with $g(\alpha) = 0$ is divisible by p.*

Proof. The first part follows, because $ev\colon F[x] \times E \to E$ is a ring homomorphism in the polynomial thus $p = q_1 q_2$ implies $0 = p(\alpha) = q_1(\alpha)q_2(\alpha)$ and since E is a field, $q_1(\alpha) = 0$ or $q_2(\alpha) = 0$. W.l.o.g. the former equality. But then $\deg q_1 = \deg p$ and so $q_2 \in F^*$ a unit and thus p irreducible.

For the second statement note that $F[x]$ is a Euclidean domain Definition 2.4.4 and thus $d := \gcd(p, g)$ has root α. If $\deg d < \deg p$, then p is not minimal, a contradiction. Otherwise $p = ud$ for some unit $u \in F[x]^* = F^*$ and therefore $p|g$. □

Lemma 3.2.4. *Given algebraic field extensions $F \subset K \subset E$ together with an F-homomorphism $\phi\colon K \to E$, then there exists an F-automorphism σ of E such that $\phi = \sigma \circ k$, where $k\colon K \hookrightarrow E$ is the embedding of K into E.*

Proof for finite extensions. If $[E : K]$ is finitely generated, e.g. $E = K(\alpha_1, \ldots, \alpha_n)$ for some $\alpha_k \in E \setminus K$, then all we have to do is extend the embedding ϕ consistently to $\alpha_1, \ldots, \alpha_n$. Consider thus a minimal polynomial $p \in F[x]$ for α_1. Obviously $p_\phi = p$, because ϕ is the identity on F. On the other hand p may break over K into smaller irreducible factors $p = p_1^{n_1} \cdots p_k^{n_k}$ one of which has root α_1, w.l.o.g. p_1. Then ϕ has to map α_1 to a root of $(p_1)_\phi$ which has the same degree as p_1 and therefore at least one zero. This extends ϕ to an embedding $\tilde{\phi}\colon K(\alpha_1) \to E$. In order to further extend it to E we go inductively over the remaining elements $\alpha_2, \ldots, \alpha_n$. In the last step we obtain an embedding of E into itself which therefore must be an isomorphism. □

For extensions of infinite degree you need something like Zorn's lemma that says that there is a limiting construction $\tilde{\phi}$ that extends to all of E. Details can be found in Appendix B.

Definition 3.2.5. *A field F is called* algebraically closed (代数闭包的) *if every polynomial over F of degree at least 1 has a root in F.*

Example 3.2.6 (Fundamental theorem of algebra)**.** Remember from complex analysis that every complex polynomial of degree at least 1 has a complex root. Therefore $\mathbb{C}$ is an algebraically closed field. It means in particular that every polynomial factors uniquely into a product of linear terms where the uniqueness is up to rearrangement and a factor $u \in \mathbb{C}^*$ which can be made unique by requiring $p = p_n(x - x_1)\cdots(x - x_n)$ with p_n the leading coefficient of p and $x_1, \ldots, x_n \in \mathbb{C}$ the roots of p.

The rational numbers $\mathbb{Q}$ on the other hand, have plenty of non-trivial irreducible polynomials, e.g. $x^2 - 2$, because its roots are $\pm\sqrt{2} \notin \mathbb{Q}$. The question is now whether we can make every field algebraically closed by adding some numbers. The general idea is the following.

Lemma 3.2.7. *Given an irreducible polynomial $p \in F[x]$, then the extension field $E := F[\xi]/(p)$ has a root of p.*

Proof. The obvious root would be ξ, but we first need to make sure, that E is a field. Certainly $F[\xi] \cong F[x]$ is a (the polynomial) ring over F and in particular a principal ideal domain. Moreover $(p) \triangleleft F[\xi]$ is a principal ideal. Since p is irreducible, (p) is a prime ideal (Proposition 2.3.6-4). But then $(p) \triangleleft F[x]$ is also a maximal ideal (Proposition 2.4.7) and thus $F[\xi]/(p)$ a field that contains F as a subfield. □

In this sense we can try to construct the algebraic closure of F by adding more and more roots of (irreducible) polynomials over F, but the full proof requires Zorn's lemma.

Theorem 3.2.8. *Let F be a field. It has an* algebraic closure*, i.e. an algebraic field extension $F \subset \bar{F}$ where $\bar{F}$ is algebraically closed. $\bar{F}$ is unique up to isomorphism of extensions.*

Easier part of the proof. We assume that there is an extension $F \subset E$ that contains all algebraic elements over F, i.e. for every non-constant polynomial $p \in F[x]$ there is at least one (and thus $\deg p$) element(s) $\alpha \in E$ such that $p(\alpha) = 0$. The difficulty is that E may be too big, i.e. also contain transcendental elements over F. We thus take the set $\mathrm{Alg}(E/F) := \{\alpha \in E : \alpha \text{ is algebraic over } F\} \subset E$ and define $\bar{F} := F(\mathrm{Alg})$. Clearly $\bar{F}$ is algebraic over F. Let $p \in \bar{F}[x]$ be any non-constant polynomial over $\bar{F}$. Since it has only finitely many coefficients, all these coefficients are contained in a finite (algebraic) extension $F \subset F_p \subset \bar{F}$. Now there is a field $E_p \supset F_p$ of finite degree that contains a root of $p \in F_p[x]$. But since $[F_p : F]$ as well as $[E_p : F_p]$ are finite, so is $[E_p : F] = [E_p : F_p][F_p : F]$ and

therefore every element in E_p including any root of p is algebraic over F. Thus there is a root of p in $\bar{F}$.

In this way there is an embedding of every algebraic element α over F in any algebraic closure of F. But this embedding can be made into a ring-homomorphism that extends the isomorphism of the base field F and thus it is an embedding homomorphism with an inverse and therefore an isomorphism of any two algebraic closures of F. □

The part about the existence of any such E is postponed to Appendix B as it needs a better understanding of Zorn's lemma (which is not in the focus of this course).

Example 3.2.9.

1. The algebraic numbers $\bar{\mathbb{Q}}$ are all complex numbers that are algebraic over $\mathbb{Q}$. It is an algebraically closed field, the algebraic closure of $\mathbb{Q}$. Note that this is strictly smaller than $\mathbb{C}$, because of transcendental numbers such as e and π.

2. Consider the real numbers $\mathbb{R}$. We know that the only irreducible polynomials over $\mathbb{R}$ are quadratic polynomials $x^2 + px + q$ with negative discriminant $D := p^2 - 4q < 0$. But these all have complex roots. Therefore $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$.

The isomorphism of all algebraic closures of any fixed field F also implies the following:

Corollary 3.2.10. *Given any algebraic field extension $f_0 \colon F \hookrightarrow E$ and any F-homomorphism $\phi \colon E \to \bar{F}$, i.e. $\phi \circ f_0 = (\bar{f} \colon F \hookrightarrow \bar{F})$, then there is an F-automorphism $\sigma \in \mathrm{Aut}_F(\bar{F})$ such that $\phi = \sigma \circ e_0$ with $e_0 \colon E \hookrightarrow \bar{E} = \bar{F}$.* □

In the following lectures we will thus assume that for every given field F there is always a smallest field $\mathbb{F}_p \subset F$ ($p \in \mathbb{P}$ a prime or $p = 0$ and $\mathbb{F}_0 := \mathbb{Q}$) as well as a fixed field $\bar{F}$ into which all the algebraic elements over F are embedded.

The Lemma 3.2.7 also permits us to construct for every polynomial $p \in F[x]$ over a field F an algebraic extension $F \subset E_p$ in which p factors into linear terms. Given an algebraic closure $\bar{F}$ we just pick all roots $\alpha_1, \ldots, \alpha_n$ of p in $\bar{F}$ and define $E_p := F(\alpha_1, \ldots, \alpha_n)$ which is a finite extension. This is called the *splitting field (分裂域)* of p.

3.2.99 Exercises

Exercise 3.2.1. Suppose that a, b are algebraic over the field F (with minimal polynomials) of degree m and n, respectively. What can you say about the degree (of the minimal polynomials) of $a \pm b$, ab, a/b (for $b \neq 0$)?

Exercise 3.2.2. Show that every algebraically closed field is infinite.
Hint: Euclid's theorem.

3.3 Separable extensions (可分扩张)

Definition 3.3.1. *An irreducible polynomial $p \in F[x]$ is* separable (可分) *if it has no multiple roots in $\bar{F}$.*[1]

An element $\alpha \in E$ of a field extension $F \subset E \subset \bar{F}$ is called separable if there is a nontrivial polynomial $p \in F[x]$ over F that has only simple roots in $\bar{F}$ and $p(\alpha) = 0$.

An algebraic field extension $F \subset E$ is called separable iff every element $\alpha \in E$ is separable over F.

Proposition 3.3.2. *For fields of characteristic 0 and the prime fields $\mathbb{F}_p$ with $p \in \mathbb{P}$ a prime all algebraic extensions are separable.*

Idea of proof. Suppose $\alpha \in \bar{F}$ is algebraic over F. Then there is a minimal polynomial $p \in F[x]$. If α is a multiple root of p, then p factors over $\bar{F}$ as $p = (x-\alpha)^n p_1$ with some $p_1 \in \bar{F}[x]$. In characteristic 0 we know that the standard derivative $\partial\colon \bar{F}[x] \to \bar{F}[x] : \bar{F} \to 0, x \mapsto 1$ fulfills $\partial p = 0$ iff $p \in \bar{F}$. As showed in Exercise 2.3.4, the greatest common divisor $d := \gcd(p, \partial p) \in F[x]$ of p and its derivative has roots in $\bar{F}$ exactly the repeated roots of p. If d is not a unit, then $p_2 := p/d \in F[x]$ is a polynomial with root α in $\bar{F}$ of lower degree than p. This is in contradiction to the assumption that p was minimal.

Consider now a finite dimensional algebraic extension $\mathbb{F}_p(\alpha)$ of some $\mathbb{F}_p$. Since $\alpha \in \bar{\mathbb{F}}_p$ is algebraic there is a corresponding minimal polynomial q and we can correspondingly define the derivative $\partial\colon \bar{\mathbb{F}}_p[x] \to \bar{\mathbb{F}}_p[x] : \bar{\mathbb{F}}_p \to 0, x \mapsto 1$. Note that there are non-constant polynomials $f \in \mathbb{F}[x]$ with $\partial f = 0$, because $\mathbb{F}_p$ does not have characteristic 0.[2] Nevertheless α is a root of ∂f iff it is a multiple root of f. But then again $d := \gcd(q, \partial q) \in F[x]$ has zeros in $\bar{F}$ the multiple roots of q. So **MG:** ? q/d would be a non-trivial polynomial with root α of degree strictly less than that of f in contradiction to q being a minimal polynomial. □

Example 3.3.3. As an example of a non-separable field extension, consider the transcendental field extension $F := \mathbb{F}_2(\xi)$ and the irreducible polynomial $f =$

[1] There are two generalizations to arbitrary polynomials: the generally accepted (1) an arbitrary polynomial is separable iff all its irreducible factors are separable; or the more naive (2) an arbitrary polynomial is separable iff it has no multiple roots in the splitting field, which is claimed, e.g. in our textbook. While (1) is true for arbitrary polynomials that have separable roots, the stronger claim (2) is true for the irreducible polynomial of a separable element even when lifted to an algebraic extension (where it may break up into several irreducible factors).

[2] Find one!

$t^2 - \xi \in F[t]$. Obviously in the field $F[\eta]/(f)$ it splits into linear factors, but it splits as $f = (t-\eta)^2$ and therefore is not separable.

It turns out that the different roots of an irreducible polynomial $p \in F[x]$ lead to F-automorphisms of the splitting field $E_p \supset F$ as follows.

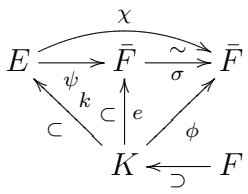
Proposition 3.3.4. *Given an algebraic element α over F. If its minimal polynomial has s different roots in the algebraic closure $\bar{F}$, then there are exactly s different F-homomorphisms of $F(\alpha)$ into $\bar{F}$.*

Proof. Since $\alpha \in \bar{F}$ as well as $F \subset \bar{F}$ there is at least one homomorphism of $F(\alpha)$ into $\bar{F}$. Let $\alpha = \alpha_1, \ldots, \alpha_s$ denote the different roots of p in $\bar{F}$. Since p is irreducible it is the minimal polynomial for every α_k, the map $\phi_k \colon F(\alpha) \to \bar{F} : F \to F \subset \bar{F}, \alpha \mapsto \alpha_k$ extends uniquely to a ring homomorphism. But there are exactly s such maps. Conversely given any homomorphism $\phi \colon F(\alpha) \to \bar{F}$ that leaves F invariant, then α must be mapped to a root of p, hence $\phi = \phi_k$ for some $1 \leq k \leq s$. □

Definition 3.3.5. *Given an algebraic field extension $F \subset E$, then we denote $[E : F]_s$ the number of F-homomorphisms of E into $\bar{F}$, called the* separability degree (可分度数).

We have just shown above that $[F(\alpha) : F]_s$ is the number of different roots of the minimal polynomial of α.

Lemma 3.3.6 (Tower property)**.** *Given an algebraic field extension $F \subset K \subset E$, then $[E : F]_s = [E : K]_s [K : F]_s$.*



Proof. Without loss of generality we assume $E \subset \bar{F}$. Let $\phi \colon K \to \bar{F}$ be an F-homomorphism. By Corollary 3.2.10 there exists an extension $\sigma \in \mathrm{Aut}_F(\bar{F})$ such that $\phi = \sigma \circ e$. For every $\psi \in \mathrm{Hom}_K(E, \bar{F})$, i.e. $\psi \circ k = e$, the homomorphism $\chi := \sigma \circ \psi : E \to \bar{F}$ is an F-homomorphism. Moreover $\chi \circ k = \sigma \circ \psi \circ k = \sigma \circ e = \phi$, i.e. χ extends ϕ.

Conversely for every $\chi \in \mathrm{Hom}_F(E, \bar{F})$ that extends ϕ, i.e. $\chi \circ k = \phi$, then $\psi := \sigma^{-1} \circ \chi$ fulfills $\psi \circ k = \sigma^{-1} \circ \chi \circ k = \sigma^{-1} \circ \phi = e$, i.e. $\psi \in \mathrm{Hom}_K(E, \bar{F})$.

Therefore there are exactly $[E : K]_s$ F-homomorphisms of E into $\bar{F}$ that extend a particular given ϕ. Conversely we can partition the $\mathrm{Hom}_F(E, \bar{F})/\sim$

with $\chi \sim \chi'$ iff $\chi \circ k = \chi' \circ k$, i.e. their restrictions to K coincide. There are $|\operatorname{Hom}_F(K, \bar{F})| = [K : F]_s$ such classes. As before each class has $[E : K]_s$ elements. So in total there are $[E : F]_s = |\operatorname{Hom}_F(E, \bar{F})| = [E : K]_s[K : F]_s$ elements. □

Proposition 3.3.7. *Given an algebraic extension $F \subset E$, then $[E : F]_s \leq [E : F]$. In particular in every tower of algebraic extensions $F \subset K \subset E$, E/F is separable iff E/K and K/F are separable.*

Proof. If $E = F(\alpha)$, then $[E : F]_s$ can be at most $\deg p = [E : F]$ where p is the minimal polynomial of α over F. If you need more than one element two write $E = F(\alpha_1, \alpha_2, \dots)$, then the inequality is still true in every step. Together with the Tower Property this ensures the inequality. In particular if $[E : F]_s$ is infinite, then E cannot be generated by a single algebraic element over F.

In a tower of finite extensions we have $[E : K]_s[K : F]_s = [E : F]_s \leq [E : F] = [E : K]\,[K : F]$ and $[E : K]_s \leq [E : K]$, $[K : F]_s \leq [K : F]$ and thus equality in the first relation iff we have equality in the other two relations. This also implies separability for every element $\alpha \in E$ that is algebraic over F. Therefore the separability condition also holds for infinite algebraic extensions. □

Proposition 3.3.8 (Theorem of a primitive element, 本原元定理, E. Artin[3])**.** *Let $F \subset E$ be a finite separable extension, then there is an algebraic element $\alpha \in E$ such that $E = F(\alpha)$.*

Proof. If F is finite, then so is E and thus its multiplicative group E^*. But then E^* is cyclic and hence has a generator $\alpha \in E^*$.

If F is infinite, we need to show that every extension $E := F(\alpha, \beta)$ by two algebraic elements is generated by one (algebraic) element. Let $n := [E : F] = [E : F]_s$ and $\phi_1, \dots, \phi_n$ the F-homomorphisms of E into $\bar{F}$. Let further

$$p := \prod_{i<j} (\phi_i(\alpha) + \phi_i(\beta)x - \phi_j(\alpha) - \phi_j(\beta)x) \in \bar{F}[x].$$

Since F is infinite, we cannot have $p(t) \equiv 0$ for all $t \in F$. Let $t \in F$ be such that $p(t) \neq 0$. But then $\phi_1(\alpha + t\beta), \dots, \phi_n(\alpha + t\beta)$ must all be distinct. Hence there are at least n F-homomorphisms of $F(\alpha + t\beta) \subset E$ into $\bar{F}$. But then $[F(\alpha + t\beta) : F]_s \geq [E : F]_s$ and so $F(\alpha + t\beta) = E$ as required. □

Corollary 3.3.9. *Let $F \subset E$ be a separable algebraic extension and for every $\alpha \in E$, $\operatorname{Irr}(\alpha/F)$ have degree at most n. Then E/F is finite and $[E : F] \leq n$.*

[3]Emil Artin *3/1898 in Germany, †12/1962

Proof. Let $\alpha \in E$ be such that $m := \deg \mathrm{Irr}(\alpha/F)$ is maximal. For every $\beta \in E$, we have $F(\alpha, \beta) = F(\gamma)$ for some $\gamma \in E$. But then $\deg \mathrm{Irr}(\gamma/F) \leq m$ and so $[F(\gamma) : F] \leq m$. Since $F(\alpha) \subset F(\gamma)$ and $[F(\alpha) : F] = m$, we have $F(\gamma) = F(\alpha)$ and so $\beta \in F(\alpha)$, i.e. $F(\alpha) = E$ and in particular $[E : F] = [F(\alpha) : F] = m \leq n$. □

3.3.99 Exercises

Exercise 3.3.1. Let $F \subset E$ be a separable extension and K be a field with the same characteristic. Show that EK is separable over FK.

Exercise 3.3.2. Let $F \subset E$ be any algebraic extension. Show that the set $\mathrm{Sep}(E/F) := \{\alpha \in E : \alpha \text{ separable over } F\}$ is a subfield of E.

Exercise 3.3.3. Let F be a field of characteristic $p \in \mathbb{P}$ and define $F^{1/p^\infty} := \{\alpha \in \bar{F} : \exists m \in \mathbb{N} : \alpha^{p^m} \in F\}$. Show that F^{1/p^∞} is a *purely inseparable field extension (纯不可分扩张)* over F, i.e. every $\alpha \in F^{1/p^\infty}$ that is separable over F is in F.
Hint: You also have to show that F^{1/p^∞} is a field.

Exercise 3.3.4. Prove the following tower property of purely inseparable field extensions: Given a tower of algebraic field extensions $F \subset K \subset E$, then E/F is purely inseparable iff E/K and K/F are purely inseparable.

Exercise 3.3.5. Show the following properties of the *inseparability degree (纯不可分度数)* of an algebraic extension $F \subset E$ of finite degree $[E : F]_i := [E : F]/[E : F]_s$:

a. $[E : F]_i \in \mathbb{N}_+$,

b. if $[E : F]_i \geq 2$, then there is at least one purely inseparable element $\alpha \in E$;

c. if $F \subset K \subset E$ are finite algebraic extensions, then $[E : F]_i = [E : K]_i[K : F]_i$.

3.4 Resultants (结式) and discriminants (判别式)

Example 3.4.1. 2. Consider the quadratic polynomial $x^2 + px + q \in \mathbb{Q}[x]$. Its roots are $x_{1/2} = -p/2 \pm \sqrt{(\frac{p}{2})^2 - q} \in \mathbb{C}$ which are either distinct real numbers, a double solution, or conjugate complex numbers. The distinction can be made by comparing $D := p^2 - 4q$ with 0. This is called a *discriminant (判别式)*.

Before we can give general formulas for discriminants of arbitrary polynomials, we need to understand symmetric functions better.

Definition 3.4.2. *Given a field F and n indeterminates $x_1, \dots, x_n$. Then a polynomial $p \in F[x_1, \dots, x_n]$ is called* symmetric (对称多项式) *if it does not change under permutation of the variables.*

Proposition 3.4.3 (Vieta[4])**.** *Given a monic polynomial $p \in F[x]$ of degree n over a field, then the coefficients are symmetric functions in the roots $\alpha_1, \dots, \alpha_n \in \bar{F}$, namely*

$$\begin{aligned} a_{n-1} &= -\sum_i \alpha_i \\ a_{n-2} &= \sum_{i<j} \alpha_i \alpha_j \\ &\vdots \\ a_0 &= (-1)^n \prod_i \alpha_i \end{aligned}$$

Proof. In the algebraic closure F, p has a root. Via long polynomial division, p factors into a linear factor and a polynomial of one degree smaller. By induction p thus factors into n linear factors, i.e.

$$p = (x - \alpha_1) \dots (x - \alpha_n).$$

Multiplying out these linear factors we obtain the formulas of the proposition. □

Lemma 3.4.4. *Given a symmetric polynomial $p \in F[x_1, \dots, x_n]$, then this is a polynomial in the elementary symmetric functions $s_1 := \sum_i x_i$, $s_2 := \sum_{i<j} x_i x_j$, …, $s_n := \prod_i x_i$.*

Proof. Obviously $F[s_1, \dots, s_n] \subset F[x_1, \dots, x_n]^{S_n}$.

Conversely given any symmetric polynomial $p \in F[x_1, \dots, x_n]$, we can first split it into homogeneous components $p = \sum_{k=0}^{d} p_k$ where d is the degree of p and p_k are homogeneous of degree k. Then each p_k is symmetric by itself. In degree 0 we have $p_0 \in F$. In degree 1 we obtain $c_1 := p_1(1, 0, \dots, 0) \in F$ which together with linearity implies $p_1 = c_1 s_1$. In degree 2 we obtain $c_2 := p_2(1, 0, \dots, 0) \in F$, $c_{1,1} := p_2(1, 1, 0, \dots, 0) - 2p_2(1, 0, \dots, 0)$ and thus from homogeneity $p_2 = c_2 \sum_i x_i^2 + c_{1,1} \sum_{i<j} x_i x_j$. It remains to express $\sum_i x_i^2$ in terms of elementary symmetric

[4]François Viète *1540 in France, †2/1603

functions s_1 and s_2 which is left as an exercise. Correspondingly in degree d we obtain

$$p_d = \sum_{|s|=d} c_s \sum_{|k|=s} x^k$$

where $k = (k_1, k_2, \ldots, k_n) \in \mathbb{N}^n$, $x^k := x_1^{k_1} \cdots x_n^{k_n}$, $|k| = \{(k_i, r) : k_i \neq 0, r$ number of j such that $k_i = k_j\}$, and $|\{(k_i, r_i)\}| = \sum k_i r_i$ where the sum runs over the different nonzero k_i. The possible s for $d = 0$ are $(0, 0)$, for $d = 1$ are $(1, 1)$, $d = 2$ are $\{(2, 1), (1, 2)\}$, $d = 3$ are $\{(3, 1), (1, 3), \{(2, 1), (1, 1)\}\}$, ... □

Proposition 3.4.5. *Given a monic polynomial $p \in F[x]$ of degree n, then its* discriminant (判别式) *is the element*

$$D_p := \prod_{i<j} (\alpha_i - \alpha_j)^2$$

where α_i are the n roots in the splitting field/ algebraic closure $\bar{F}$. The discriminant is an element of F and 0 iff p has a multiple root.

Proof. Note that the construction shows that D_p is symmetric in all its roots. On the other hand the coefficients a_0, ..., a_{n-1} are the elementary symmetric functions in the roots and each an element of F. Therefore also the discriminant is a polynomial in the coefficients a_i and thus an element of F.

From the definition it is also clear that $D_p = 0$ iff one of the factors vanishes. But then there are two coinciding $\alpha_i = \alpha_j$ with $i \neq j$. □

Example 3.4.6. 3. (del Ferro[5], N. Tartaglia[6]) Consider the cubic polynomial $x^3 + rx^2 + sx + t \in \mathbb{Q}[x]$. The first step in finding its roots is the substitution $z = x - r/3$ which leads to the reduced polynomial $z^3 + pz + q \in \mathbb{Q}[z]$. With the substitution $z = u + v$ one obtains $u^3 + v^3 + (3uv + p)(u + v) + q = 0$ where we can impose one constraint, e.g. $uv = -p/3$, and thus obtain $u^3 + v^3 = -q$ and $u^3 v^3 = -p^3/27$ and thus u, v are the two roots of the equation $\zeta^2 + q\zeta - p^3/27 = 0$ (the *resolvent equation (预解方程)*). The discriminant is thus proportional to $D := q^2/4 + p^3/27$. The case $D > 0$ gives the real solution $u^3 = -q/2 + \sqrt{D}$ and $v^3 = -q/2 - \sqrt{D}$ ($z = u + v$ and $x = z + r/3$). The two conjugate complex solutions are $u_{2,3} = \omega_3^{1,2} u_1$ and $v_{2,3} = \omega_3^{2,1} v_1$ (and thus $z_k = u_k + v_k$ and $x_k = z_k + r/3$, where $\omega_3 := \mathrm{e}^{\pm 2\pi \mathrm{i}/3}$ is a primitive third root of 1). The interesting case is $D < 0$ (*casus irreducibilis*), because it implies that one has to use complex numbers in order to obtain 3 real roots. With the choice $u_k = \left(-q/2 - i\sqrt{-D}\right)^{1/3} \omega_3^k$, $z_k = u_k - p/(3u_k)$ (and $x = z + r/3$).

[5]Scipione del Ferro *2/1465 in Bologna/Italy, †11/1526

[6]Niccolò F. Tartaglia *1499/1500 in Brescia/Rep. of Venice (Italy), †12/1557

4. (Ferrari's[7] solution) Consider the quartic polynomial $x^4+ax^3+bx^2+cx+d$ and do the analog substitution $z = x-a/4$ that leads to $z^4+rz^2+sz+t = 0$. Adding y^2 on both sides, we complete the square to obtain $(z^2+r+y)^2 = (r+2y)z^2 - sz + (y^2+2ry+r^2-t)$. If we want the right hand side to be the perfect square $(\alpha z+\beta)^2$, then we need $2y^3+5ry^2+(4r^2-2t)y+(r^3-rt-s^2/4) = 0$. This is the resolvent of the quartic equation. With the previous method for solving cubic equations, we obtain that the discriminant is proportional to $D = Q^2/4+P^3/27$ for the coefficients of the reduced cubic form $v^3+Pv+Q = 0$, i.e. $P = -r^2/12-t$, $Q = -r^3/108+rt/3-s^2/8$.

5. **Q:** Why is there no solution formula for $x^5+ax^3+bx^2+cx+d = 0$?

Remark 3.4.7. A similar question is whether two polynomials $p, q \in F[x]$ over the same field F have common roots. The quantity to decide that is called resultant. However there is also an easier way to decide that: In the algebraic closure $\bar{F}$ both polynomials factor as $p = p_m(x-\alpha_1)\cdots(x-\alpha_m)$ and $q = q_n(x-\beta_1)\cdots(x-\beta_n)$ respectively. Now their greatest common divisor contains exactly the product of the common factors, i.e. $\gcd(p,q) = c(x-\gamma_1)\cdots(x-\gamma_r)$ where $\gamma_i = \alpha_j = \beta_k$ for increasing sequences $j\colon \{1,\dots,r\} \to \{1,\dots,m\}$ and $k\colon \{1,\dots,r\} \to \{1,\dots,n\}$. But the gcd can also be computed from the coefficients of p and q without knowing their roots via Euclid's algorithm. Moreover the degree of the gcd does not only tell you if there are common roots ($\deg > 0$), but also how many.

3.4.99 Exercises

Exercise 3.4.1. When do x^2+ax+b and $x^2+px+q \in F[x]$ have common roots?

Exercise 3.4.2. Verify the formula for the discriminant of x^4+rx^2+sx+t.

Exercise 3.4.3. Write the symmetric function $p_d(x_1,\dots,x_n) := x_1^d+\cdots+x_n^d \in F[x_1,\dots,x_n]$ as a polynomial $\bar{f} \in F[s_1,\dots,s_n]$ in the elementary symmetric functions

a. for $d = 2$,

b. for $d = 3$,

c. for arbitrary $d \in \mathbb{N}$.

Exercise 3.4.4. We know that for quadratic polynomials $x^2+px+q \in \mathbb{R}[x]$ the polynomial factors over $\mathbb{R}$ iff the discriminant $D := p^2-4q$ fulfills $D \geq 0$. What is the corresponding condition for quadratic polynomials over arbitrary fields F?

[7] Lodovico Ferrari *2/1522 in Milan/Italy, †10/1565

3.5 Splitting fields and Normal extensions (正规扩张)

Remember that the splitting field E_p of a polynomial $p \in F[x]$ over a field F is the algebraic extension of F by all roots of p in $\bar{F}$.

These splitting fields have an interesting property:

Proposition 3.5.1. *Let $F \subset E$ be the splitting field of a polynomial $p \in F[x]$, then*

1. *every F-homomorphism $\phi\colon E \to \bar{F}$ has $\phi(E) = E$,*
2. *every irreducible polynomial over F that has a root in E splits into linear factors over E.*

Proof. "1" Note that every F-homomorphism leaves F invariant and thus maps p to itself and therefore also all its roots in E to roots in $\bar{F}$ of the same p. But all these roots are in E. Since E is generated by all these roots, $\phi(E) = E$.

"2" Suppose thus that $q \in F[x]$ is an irreducible polynomial and has a root $\alpha \in E$. But due to Proposition 3.3.4, there are F-automorphisms of $\bar{F}$ that map the root α to every other root of q. Each of those automorphisms restricts to a homomorphism of E into $\bar{F}$ and by the first part map $\phi(E) = E$. But then every other root of q is also in E, i.e. q splits into linear factors over E. □

Example 3.5.2. Note however, that this is not true for arbitrary field extensions. Let $E := \mathbb{Q}(\sqrt[3]{2})$ and $p = x^3 - 2 \in \mathbb{Q}[x]$ which is an irreducible polynomial over $\mathbb{Q}$. It has a root in E, namely $\sqrt[3]{2}$, but splits here only into $p = (x - \sqrt[3]{2})(x^2 + \sqrt[3]{2}x + \sqrt[3]{4})$ and the latter is irreducible over E, because its roots $\mathrm{e}^{\pm 2i\pi/3}\sqrt[3]{2}$ are not in E.

We therefore define.

Definition 3.5.3. *A field extension $F \subset E$ is called* normal (正规的) *if every irreducible polynomial in F that has a root in E splits into linear factors over E.*

Actually the first proposition is an equivalence if we either restrict to finite dimensional extensions or consider splitting fields of arbitrary families of polynomials over F.

A property we will need for the Galois correspondence principle is the following:

Lemma 3.5.4 (Normal tower)**.** *Given a normal field extension $F \subset E$, then for every intermediate field $F \subset K \subset E$ the extension $K \subset E$ is also normal.*

Proof. If E is the splitting field of the (family of) polynomial(s) $p \in F[x]$, then it is also the splitting field of the same polynomials over K. □

Remark 3.5.5. Note however that the other three cases can be wrong, i.e. neither K/F need to be normal nor does K/F and E/K being normal imply that E/F is normal.

Remark 3.5.6. Analogous to normal subgroups it is also possible to define the normal closure of an intermediate field $F \subset K \subset \bar{F}$ as the smallest normal extension of F in $\bar{F}$ that contains K, because the intersection of normal extensions is normal again. Correspondingly this normal closure is generated (as the composite field) by all conjugates of K in $\bar{F}/F$, i.e. the image of K under F-automorphisms of $\bar{F}$.

3.5.99 Exercises

Exercise 3.5.1. Find counter examples for the Remark 3.5.5, i.e.

0 a normal field extension $F \subset E$ together with an intermediate field $F \subset K \subset E$ such that K/F is not normal;

a. two normal field extensions $F \subset K$ and $K \subset E$ such that E/F is not normal.

Exercise 3.5.2 (Structure of finite fields)**.** Show that $\mathbb{F}_{p^m}$ and $\mathbb{F}_{p^n}$ are embedded in $\mathbb{F}_{p^l}$ with $l = \operatorname{lcm}(m, n)$ and their intersection (in the embedding) is $\mathbb{F}_{p^d}$ with $d = \gcd(m, n)$.

Conclude that $\bar{\mathbb{F}}_p$ is the inductive limit $\bar{\mathbb{F}}_p = \varinjlim_n \mathbb{F}_{p^n}$, what are the embeddings $\mathbb{F}_{p^m} \hookrightarrow \mathbb{F}_{p^n}$ (i.e. for which m and n do they exist)?

Exercise 3.5.3. Consider a field $F \subset \bar{F}$ together with a family of intermediate fields $F \subset E_\alpha \subset \bar{F}$ and prove the following:

a. if all E_α are normal over F, then so is their intersection;

b. the normal closure of an algebraic extension $F \subset E \subset \bar{F}$ is the composite of all conjugates of E, i.e. the images of E under all F-automorphisms of $\bar{F}$.

3.6 Galois extensions (伽罗瓦扩张) and the correspondence principle (对应原理)

Lemma 3.6.1. *$\mathbb{Q}$ and the prime fields $\mathbb{F}_p$ for $p \in \mathbb{P}$ a prime have no automorphisms.*

Proof. The reason is that $\mathbb{Q}$ and $\mathbb{F}_p$ are the smallest fields generated only by 1. Namely let $\phi\colon \mathbb{Q} \to \mathbb{Q}$ be a ring homomorphism. In particular $\phi(1) = 1$. But then

$\phi(n\alpha) = n\phi(\alpha)$ by induction. But this implies that $\phi(p/q) = p/q$ and thus ϕ is the identity on $\mathbb{Q}$.

The arguments for the other fields are analogous. □

Example 3.6.2. The situation is completely different for the complex numbers $\mathbb{C}$. Consider, e.g. complex conjugation $\bar{}\colon \mathbb{C} \to \mathbb{C} : a + b\mathsf{i} \mapsto a - b\mathsf{i}$ for $a, b \in \mathbb{R}$. It is easy to verify that this is a ring homomorphism. Obviously complex conjugation is not the identity on $\mathbb{C}$.

Proposition 3.6.3. *Given the splitting field $F \subset E$ of an irreducible separable polynomial $p \in F[x]$ of degree n. Then $\mathrm{Aut}_F(E)$ is a transitive subgroup of S_n.*

Proof. Obviously automorphisms of E are fixed by specifying the image of all the roots $\xi \in E$ of p. But due to the automorphism property these can only be mapped to other roots of p. Since p is separable it has exactly $n = \deg p$ different roots and any automorphism must thus be a permutation of these roots. This implies $\mathrm{Aut}_F(E) \subset S_n$.

Let now ξ_1 be one of the roots and ξ_2 another one. Consider the extension $K := F(\xi_1, \xi_2)$ obviously $F \subset K \subset E$. Since each is finite dimensional, they are both algebraic extensions. Since ξ_i have both the same minimal polynomial p, we can construct an automorphism $\mathrm{Aut}_F(K) \ni \sigma\colon \xi_1 \mapsto \xi_2$. We can extend this automorphism from K to E and thus obtain an automorphism $\tilde{\sigma} \in \mathrm{Aut}_F(E)$ that maps ξ_1 to ξ_2, for every pair of roots ξ_k of p. This completes the proof. □

Warning: This does not imply that the automorphism group is S_n, but rather a transitive subgroup, e.g. C_n, A_n (for $n \geq 3$), D_n (for $n \geq 3$), or S_n.

Definition 3.6.4. *A field extension $F \subset E$ is called* Galois (伽罗瓦的) *if it is separable and normal.*

The Galois group of an algebraic extension $F \subset E$ is $\mathrm{Gal}(E : F) := \mathrm{Aut}_F(E)$ *the set of all automorphisms of E that leave F invariant.*

Remark 3.6.5. Remember the definition of separability degree of an (algebraic) extension $F \subset E$ as the number of F-homomorphisms of E into $\bar{F}$. If E/F is normal (as in a Galois extension), then each such homomorphism sends E to itself, thus $|\mathrm{Aut}_F(E)| = [E : F]_s$. If moreover E is separable over F, then $[E : F]_s = [E : F]$. In total we obtain $|\mathrm{Gal}(E : F)| = [E : F]$.

The importance of Galois extensions comes from the following correspondence principle.

Theorem 3.6.6 (Galois[8]). *Given a Galois extension $F \subset E$, then there is a 1:1-correspondence between intermediate fields $F \subset K \subset E$ and subgroups $H \subset \mathrm{Gal}(E:F)$, via*

$$H \mapsto E^H := \{\alpha \in E : \forall \sigma \in H : \sigma(\alpha) = \alpha\}, \tag{3.1}$$
$$K \mapsto \mathrm{Gal}(E:K) \tag{3.2}$$

Moreover, Galois extensions $F \subset K$ with $K \subset E$ correspond to normal subgroups $H \lhd \mathrm{Gal}(E:F)$ and $\mathrm{Gal}(K:F) \cong \mathrm{Gal}(E:F)/\mathrm{Gal}(E:K)$.

Proof. From Proposition 3.3.7 and Lemma 3.5.4 it follows that E/K is always a Galois extension. Note that E^H is a ring, because the elements of $\mathrm{Gal}(E:F)$ are ring automorphisms. Also E^H contains the inverses for the same reason. Therefore E^H is a field and also contains F.

Let us verify that the operations are inverse to each other. Consider thus $k := E^{\mathrm{Gal}(E:K)}$. Obviously $K \subset k$. Assume there were an $\alpha \in k \setminus K$. But then there is an automorphism $\sigma \in \mathrm{Gal}(E:K)$ such that $\sigma(\alpha) \neq \alpha$, because α is F-linear independent from K. This would contradict k is $\mathrm{Gal}(E:K)$-invariant and therefore $k = K$.

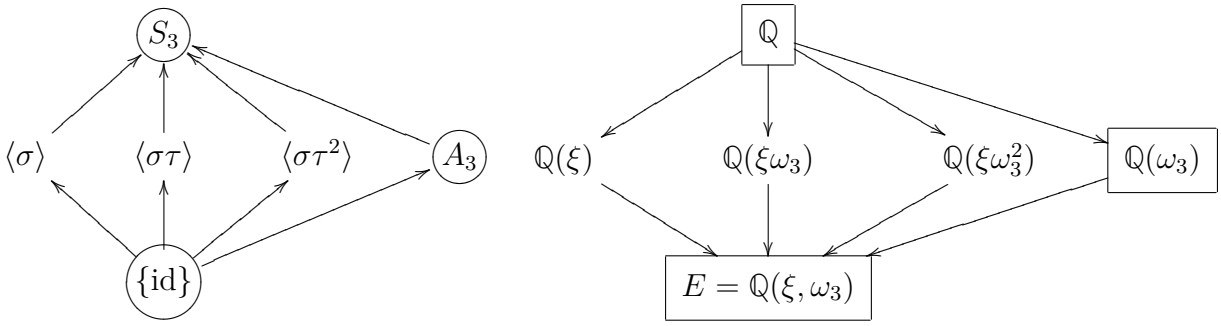
Conversely let $H' := \mathrm{Gal}(E:E^H)$. Clearly $H \subset H'$. But by Remark 3.6.5 we also have $|H| = [E:E^H] = |H'|$ both finite, so $H = H'$.

It remains to show that Galois extensions correspond to normal subgroups. Let thus $H \lhd G := \mathrm{Gal}(E/F)$ be a normal subgroup. As shown so far it corresponds to an intermediate field $F \subset E^H \subset E$. We also know that H is its own conjugate in G. That implies that $K := E^H$ is its own conjugate under all $\sigma \in G = \mathrm{Gal}(E:F)$. But then K/F is normal. This completes the proof. □

Example 3.6.7. Consider the extension $\mathbb{Q} \subset E := E_{x^3-2}$, the splitting field of $p = x^3 - 2 \in \mathbb{Q}[x]$. Beside $\xi = \sqrt[3]{2}$ is also contains the element $\omega_3 = \mathrm{e}^{2\pi\mathrm{i}/2}$ a third root of unity. Since the Galois group permutes the three solutions $\xi\omega_3^k$ of $p = x^3 - 2 \in \mathbb{Q}[x]$, it is a transitive subgroup of S_3, i.e. either A_3 or S_3 itself. Moreover the map $\sigma\colon E \to E : \xi \mapsto \xi, \omega_3 \mapsto \omega_3^{-1}$ is also an automorphism of E of order 2. Therefore $\mathrm{Gal}(p/\mathbb{Q}) = S_3$. The subgroups are $\{1, \langle\sigma\rangle, \langle\sigma\tau\rangle, \langle\sigma\tau^2\rangle, A_3, S_3\}$ and the corresponding fields $\{E, \mathbb{Q}(\xi), \mathbb{Q}(\xi\omega_3), \mathbb{Q}(\xi\omega_3^2), \mathbb{Q}(\omega_3), \mathbb{Q}\}$ and fit into the

[8] Évariste Galois *10/1811 in Bourg-la-Reine/France, †5/1832 in a duel (with guns) presumably about love but under rather dubious circumstances. Fortunately he wrote down his genial ideas before he entered the duel.

pattern:



Note that the directions of inclusion are reverted in the correspondence. The encircled subgroups are normal subgroups of S_3 and correspond to Galois extensions $K/\mathbb{Q}$ (boxed).

The last statement in the Galois correspondence is very similar to the Second Isomorphism Theorem (for groups). So one may wonder if there is any (non-trivial) correspondence to the Third Isomorphism Theorem. The answer is the following:

Proposition 3.6.8. *Given a finite Galois extension $F \subset E$ together with any field extension $F \subset K \subset L$ and also $E \subset L$, then the composite $EK \subset L$ exists, is Galois over K, as well as E is a finite Galois extension over $E \cap K$, and* $\mathrm{Gal}(EK : K) \cong \mathrm{Gal}(E : (E \cap K))$.

Proof. Since E is normal over F, it is also normal over $F \subset E \cap K$, and thus every K-automorphism of EK has a restriction to E which is an $E \cap K$-automorphism. This yields a group homomorphism $\Theta\colon \mathrm{Gal}(EK : K) \to \mathrm{Gal}(E : (E \cap K)) : \sigma \mapsto \sigma|_E$. Since EK is generated by $E \cup K$, an F-homomorphism of EK is uniquely determined by its restrictions to E and to K. Therefore Θ is injective.

If $\alpha \in E$, then $(\sigma|_E)(\alpha) = \alpha$ for all $\sigma \in \mathrm{Gal}(EK : K)$ if and only if $\sigma\alpha = \alpha$ for all $\sigma \in \mathrm{Gal}(EK : K)$, if and only if $\alpha \in K$. Thus $E \cap K$ is the fixed field of $\operatorname{im}\Theta \subset \mathrm{Gal}(E : (E \cap K))$. But then Θ must be surjective and thus an isomorphism. □

3.6.1 Galois group of polynomials of low degree

We denote $\mathrm{Gal}(p/F)$ the Galois group of the splitting field E_p of a separable polynomial $p \in F[x]$. In this way the Galois theory decides about the structure of the roots of a polynomial p.

Example 3.6.9. 2. A monic quadratic polynomial $p = x^2 + px + q \in F[x]$ over a field not of characteristic 2 is irreducible iff $D_p := p^2 - 4q$ is not a square in F. In this case the Galois group is $C_2 \cong S_2$.

Note also that the Galois group of two polynomials $p_1 p_2$ that do not have roots in common $\gcd(p_1, p_2) \in F^*$ is $\mathrm{Gal}(p_1 p_2/F) = \mathrm{Gal}(p_1/F) \times \mathrm{Gal}(p_2/F)$.

Remember the definition of the discriminant of a polynomial $p \in F[x]$ of degree n with roots $\alpha_1, \ldots, \alpha_n \in \bar{F}$ as follows

$$D_p := \prod_{i<j} (\alpha_i - \alpha_j)^2.$$

Since the discriminant is symmetric in all roots, it is a polynomial in its symmetric functions, the coefficients of the polynomial. Therefore $D_p \in F$.

In general the discriminant tells us the following about the Galois group:

Proposition 3.6.10. *Given a separable polynomial $p \in F[x]$, then the Galois group $\mathrm{Gal}(p/F)$ has an odd permutation iff its discriminant is not a square in F.*

Proof. The element $t := \prod_{i<j}(\alpha_i - \alpha_j)$ in any specific order of the roots of p is an element of the splitting field E_p of p. Moreover $D_p = t^2$ and so $t \in F(t)$ an extension of degree 1 or 2. Given any homomorphism $\sigma \in \mathrm{Gal}(p/F)$ of E_p into $\bar{F}$, then it maps $t \to \pm t$ depending on its sign in S_n. If $t \in F$, i.e. D_p has a square root in F, then no permutation can change the sign of t and so $\mathrm{Gal}(p/F) \subset A_n$. Conversely if there is any odd permutation in $\mathrm{Gal}(p/F)$, then t changes sign under this permutation and can thus not be an element of F. □

In degree 3 we get the following result:

Proposition 3.6.11. *Given a monic polynomial of degree 3, $p = x^3 + sx + t \in F[x]$ in reduced form over a field not of characteristic 2 or 3, then the Galois group $\mathrm{Gal}(p/F)$ is*

S_3 iff the discriminant is not a square in F and p irreducible, thus $[E_p : F] = 6$;

A_3 iff the discriminant is a square, but p irreducible and thus $[E_p : F] = 3$;

$\subset C_2$ iff p is reducible.

In degree 4 the result is more involved and reads

Proposition 3.6.12. *Let $p = x^4 + rx^2 + sx + t \in F[x]$ be a separable quartic polynomial in reduced form over a field not of characteristic 2. Then the order of the splitting field K of its resolvent divides 6. The Galois group of p is correspondingly*

S_4 iff $[K : F] = 6$ and p irreducible;

A_4 iff $[K : F] = 3$ and p irreducible;

D_4 iff p is irreducible over K and $[K : F] = 2$;

C_4 *iff* $[K : F] = 2$ *and* p *irreducible (over* F*), but not irreducible over* K*;*

V_4 *iff* $[K : F] = 1$ *(i.e. the resultant splits into linear factors over* F*) and* p *irreducible;*

$\subset S_3$ **or** $\subset C_2 \times C_2$ *(not transitive) iff* p *is reducible.*

The Galois group depends critically on the irreducibility (or the irreducible factors in general).

3.6.99 Exercises

Exercise 3.6.1. Let $F \subset E$ be a finite Galois extension and consider intermediate fields $F \subset K_i \subset E$ as well as corresponding subgroups $H_i \subset \mathrm{Gal}(E : F)$. Prove the following:

a. $K_1 \subset K_2$ iff $H_1 \supset H_2$;

b. $K_1 = K_2K_3$ iff $H_1 = H_2 \cap H_3$;

c. $K_1 = K_2 \cap K_3$ iff $H_1 = \langle H_2, H_3\rangle_{\mathrm{Gal}}$.

Exercise 3.6.2 (Galois connection)**.** Given two partially ordered sets $(X, \leq)$ and $(Y, \leq)$ together with order reversing maps $F\colon X \to Y$ and $G\colon Y \to X$, i.e. for all $x_i \in X$ with $x_1 \leq x_2$ then $F(x_2) \leq F(x_1)$ and for all $y_i \in Y$ with $y_1 \leq y_2$ we have $G(y_2) \leq G(y_1)$. Show that F and G induce mutually inverse order-reversing bijections between $X^{GF} := \{x \in X : (GF)(x) = x\}$ and $Y^{FG} := \{y \in Y : (FG)(y) = y\}$.

Remark 3.6.13. The pair (F, G) is called a Galois connection between X and Y.

Exercise 3.6.3. Let $p = x^3 + x - 1 \in \mathbb{Q}[x]$. Compute its Galois group together with all intermediate fields between $\mathbb{Q}$ and the splitting field E of p. Which are Galois extensions over $\mathbb{Q}$?

Exercise 3.6.4. Compute the Galois groups of the following polynomials

a. $x^3 - x - 1$ over $\mathbb{Q}(\sqrt{-23})$,

b. $x^3 - 10 \in \mathbb{Q}[x]$,

c. $x^3 - 10$ over $\mathbb{Q}(\sqrt{2})$,

d. $x^3 - 10$ over $\mathbb{Q}(\sqrt{-3})$,

e. $(x^2 - 2)(x^2 - 3) \in \mathbb{Q}[x]$,

f. $x^4 - 3 \in \mathbb{Q}[x]$

g. $x^4 - 3$ over $\mathbb{Q}(\sqrt{3})$,

h. $x^4 - 3$ over $\mathbb{Q}(\sqrt{-3})$,

i. $x^4 + x + 3 \in \mathbb{Q}[x]$,

j. $x^4 + 3x + 3 \in \mathbb{Q}[x]$.

3.7 Infinite Galois extensions (无限伽罗瓦扩张) and Picard–Vessiot theory

So far we have seen that finite dimensional extensions are algebraic. The algebraic closure of a primary field $\mathbb{Q}$ or $\mathbb{F}_p$ ($p \in \mathbb{P}$) is infinite dimensional over the primary field, but all we need in order to understand an algebraic equation (i.e. $q(x) = 0$ for some $q \in F[x]$) are the finite dimensional extensions by all the roots of the equation.

Now we consider the transcendental extensions as follows.

Definition 3.7.1. *Given a transcendental extension $F \subset E$. Its* transcendence degree (超越次数) *is the number of algebraically independent elements $\alpha_i \in E$.*

Example 3.7.2. Consider the field $\mathbb{C}(x_1, \ldots, x_n)$ of rational functions in n variables. This has transcendence degree n, because $x_1, \ldots, x_n$ are n algebraically independent elements and every other element is a rational function in them and thus the root of a polynomial in X with coefficients in $\mathbb{C}(x_1, \ldots, x_n)$.

Remember the definition of differential ∂ of a ring R, as an additive map that fulfills the Leibniz rule

$$\partial(fg) = f\partial g + g\partial f$$

Lemma 3.7.3. *The separable algebraic extensions of the prime fields $\mathbb{F}$ do not have (nontrivial) differentials.*

Proof. First, note that $1\partial 1 = \partial 1 = \partial(1 \cdot 1) = 2\partial 1$ and thus $\partial 1 = 0$. But since ∂ is additive by induction $\partial(n\alpha) = n\partial\alpha$ for $n \in \mathbb{Z}$ and $\alpha \in F$. Thus for any $\alpha = p/q$ with $p, q \in \mathbb{Z}$ and $q \neq 0 \in \mathbb{F}$, we have $\partial(p/q) = p/q\partial 1 = 0$.

Let now $\alpha \in E \supset \mathbb{F}$ be an element that is separable and algebraic over the prime field $\mathbb{F}$. It is thus a root of a separable minimal polynomial $p \in F[x]$ of degree at least 2. Starting from the equation $p(\alpha) = 0$ we derive to $0 = p'(\alpha)\partial\alpha$. Separability implies $\gcd(p, p') = 1$, but then the second factor $\partial\alpha = 0$ must vanish. Thus ∂ vanishes on the whole separable algebraic extension. □

Example 3.7.4. The interesting cases are thus transcendental extensions. Let thus $\mathbb{C} \subset E := \mathbb{C}(x)$ be the field of rational functions and define a differential by $\partial x = p$ where $p \in \mathbb{C}(x)$ is any rational function. Extending ∂ to the whole of E via the Leibniz rule and via $\partial(x^{-1}) = -(\partial x)/x^2$ we see that the differentials of $\mathbb{C}(x)$ are characterized by the element $p \in \mathbb{C}(x)$.

Definition 3.7.5. *Given a differential field (F, ∂), then a linear (homogeneous) differential equation is a polynomial $D \in F[\partial]$, the non-commutative algebra generated by the element ∂. A solution in a differential extension $F \subset (E, \partial)$ is an element $f \in E$ such that*

$$D[f] = 0.$$

Theorem 3.7.6 (Picard[9]–Lindelöf[10], Cauchy[11]–Lipschitz[12])**.** *Given an ODE as above with leading coefficient $\operatorname{lc} D = 1 \in F$ over the base field $F = \mathbb{R}(x)$ together with initial data $x_0 \in \mathbb{R}$ with $D|_{x_0} \in \mathbb{R}^{\deg D}$, $(y_0, y_1, \dots, y_{(\deg D)-1}) \in \mathbb{R}^{\deg D}$, then there is a unique solution $f \in C^{\deg D}(U)$ fulfilling the ODE and the initial data $(\partial^k f)(x_0) = y_k$ where $U \subset \mathbb{R}$ is the largest interval around x_0 where all coefficient functions of D are continuous.*

This is a standard theorem of real analysis.

Example 3.7.7. Consider the field $F = \mathbb{R}(x)$ of rational functions in one variable. Then $D := \partial^2 + \mathbb{1}$ gives a linear 2nd order ODE $y'' + y = 0$ with general solution $C_1 \cos x + C_2 \sin x \in \mathrm{C}^\omega(\mathbb{R})$.

Algebraically, if we extend (F, ∂) to $E := F(\mathrm{Sin}, \mathrm{Cos})$ with $\partial\mathrm{Sin} = \mathrm{Cos}$, $\partial\mathrm{Cos} = -\mathrm{Sin}$, then we see that E contains the general solution $C_1\mathrm{Cos} + C_2\mathrm{Sin}$ with $C_i \in \mathrm{Const}(F) = \mathbb{R}$.

Definition 3.7.8. *Given a differential field (F, ∂) together with a differential extension $F \subset (E, \partial)$ such that $\mathrm{Const}(E) = \mathrm{Const}(F)$ and there is a nontrivial linear ODE D over F of degree n that has an n-dimensional solution space over $\mathrm{Const}(E)$ and $E = F(\ker(D|E))$, then (E, ∂) is the* Picard–Vessiot[13] extension *of D over F.*

The differential Galois group $\mathrm{Gal}(E : F, \partial)$ is the group of differential field automorphisms of (E, ∂) that leave (F, ∂) invariant.

Example 3.7.9. Consider again the extension $E := F(\mathrm{Sin}, \mathrm{Cos})$ of $F := \mathbb{R}(x)$. This is a Picard–Vessiot extension, because $\ker \partial|E = \mathbb{R} = \mathrm{Const}(F)$, and $D =$

[9] Émile Picard *1856/7 in Paris/France, †1941/12
[10] E.L. Lindelöf *1870/4 in Helsinki/Finland, †1946/6
[11] Augustin-Louis Cauchy *1789/8 in Paris, France, † 1857/5
[12] Rudolph O.S. Lipschitz *1832/5 in Kalinigrad/Prussia, †1903/10
[13] Ernest Vessiot *1865/3 in Marseille/France, †1952/10

$\partial^2 + \mathbb{1}$ is of degree 2 and has a 2-dimensional solution space in E spanned by Sin and Cos.

Its differential Galois group is $\mathrm{Gal}(E : F, \partial) = \langle \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} : \alpha \in \mathbb{R}^* \rangle \cong C_2 \ltimes \mathbb{R}^*$ with the group action $\mathrm{Gal} \ni g \colon \binom{\mathrm{Sin}}{\mathrm{Cos}} \mapsto g\binom{\mathrm{Sin}}{\mathrm{Cos}}$ and $\partial \circ g = g \circ \partial$.

Theorem 3.7.10 (Picard–Vessiot, correspondence principle)**.** *Given a Picard–Vessiot extension $(F, \partial) \subset E$, then the algebraic subgroups of $\mathrm{Gal}(E : F)$ correspond bijectively to the intermediate differential fields $F \subset (K, \partial) \subset E$ via*

$$K \mapsto \mathrm{Gal}(E : K)$$
$$\mathrm{Gal}(E : F) \supset H \mapsto E^H := \{f \in E : \forall \sigma \in H : \sigma(f) = f\}.$$

In particular normal subgroups $N \triangleleft \mathrm{Gal}(E : F)$ correspond to Picard–Vessiot extensions $F \subset K$ and the Galois groups relate as $\mathrm{Gal}(K : F) \cong \mathrm{Gal}(E : F)/N$.

The proof is due to Picard and Vessiot and a modern version can be found, e.g., in [PS03].

Example 3.7.11. Given a Picard–Vessiot extension (E, ∂) of a differential field (F, ∂), then

1. an *exponential* over F is an element $e \in E$ such that $\partial e - fe = 0$ for some $f \in F$;
2. a *root* is an element $r \in E$ such that $p(r) = 0$ for some $p \in F[X]$;
3. a *quadrature* is an element $i \in E$ such that $\partial i \in F$.

Definition 3.7.12. *A differential extension $F \subset (E, \partial)$ is a* Liouville[14] extension *if every element $t \in E$ is contained in a finite tower of extensions by exponentials, roots, and quadratures.*

Lemma 3.7.13. *Given a Liouville extension $F \subset (L, \partial)$ of finite transcendence degree, then this is contained in a minimal Picard–Vessiot extension $F \subset L \subset (E, \partial)$.*

Also this proof can be found, e.g. in [PS03].

Remark. Remember that a subgroup $B \subset \mathrm{GL}_N(\mathrm{Const}(F))$ of a matrix group is solvable iff the derived series terminates in $\{\mathbb{1}\}$.

[14] Joseph Liouville *1809/3 in Saint-Omer/France, †1882/9

Remark 3.7.14. Given a matrix subgroup $G \subset \mathrm{GL}_n(\mathbb{C})$, then it breaks up into disjoint (path) connected components $\pi_0(G)$ which inherit a group structure. The connected component G^o of $\mathbb{1} \in G$ is a normal subgroup of the same dimension and $\pi_0(G) \cong G/G^o$. Consider for example $\mathrm{O}(n) \subset \mathrm{GL}_n(\mathbb{R})$, it has the subgroup of same dimension $\mathrm{SO}(n) \subset \mathrm{O}(n)$ which is the connected component of $\mathbb{1}$. The second connected component is $\begin{pmatrix} -1 & \\ & \mathbb{1} \end{pmatrix} \mathrm{SO}(n)$ (the reflections).

Theorem 3.7.15 (Kolchin[15])**.** *Given a linear ODE $D \in F[\partial]$ over a differential field (F, ∂), then this has solutions in a Liouville extension iff the Picard–Vessiot extension that solves D has a solvable connected component of* id *in the differential Galois group.*

This theorem is due to Kolchin [Kol48] and a modern proof can be found, e.g., in [PS03].

Example 3.7.16. In the example of $D = \partial^2 + \mathbb{1} \in F[\partial]$ for $F = \mathbb{R}(x)$ the Galois group was $\mathbb{Z}/(2) \ltimes \mathbb{R}^*$ which is a solvable subgroup as well as its connected component of id which is isomorphic to $\mathbb{R}^*_+$. Note that the solution can be written as exponential once the algebraic extension of the fields of constants $\mathbb{R} \subset \mathbb{C} = \mathbb{R}(\mathrm{i})$ is done.

3.7.99 Exercises

3.8 Cyclotomic (分圆), Cyclic extensions (循环扩张) and Solvability by radicals (可解用根式)

Another interesting questions which was also answered by Galois is whether we can solve a particular polynomial equation in terms of radical expressions.

Definition 3.8.1. *A primitive cyclotomic polynomial is $q := (x^p - 1)/(x - 1) \in \mathbb{Z}[x]$ for some prime $p \in \mathbb{P}$.*

A cyclotomic extension is a splitting field of an irreducible cyclotomic polynomial.

Example 3.8.2. 1. Remember that the polynomial $q = (x^n - 1)/(x - 1)$ is irreducible over $\mathbb{Q}[x]$ iff $n \in \mathbb{P}$. Consider the case $p = 3$ and thus $q = (x^3 - 1)/(x - 1) = x^2 + x + 1 \in \mathbb{Z}[x]$. Its splitting field is $\mathbb{Q}(\omega_3)$ where $\omega_3 = \mathrm{e}^{2\pi\mathrm{i}/3}$ is a primitive third root of unity. Its Galois group is therefore $\mathrm{Gal}(\mathbb{Q}(\omega_3) : \mathbb{Q}) = C_2$ a cyclic group.

[15] Ellis R. Kolchin *1916/4 in New York/USA, †1991/10

2. More generally, we can define the n-th cyclotomic polynomial as $\Phi_n := \prod_{k\in(\mathbb{Z}/(n))^*}(x-\omega_n^k) \in \mathbb{C}[x]$ where $\omega_n := \mathrm{e}^{2\pi\mathsf{i}/n}$ is a primitive n-th root and $(\mathbb{Z}/(n))^*$ are the generators of $\mathbb{Z}/(n)$, thus ω_n^k the primitive n-th root. For $p \in \mathbb{P}$ we obtain $\Phi_p = (x^p-1)/(x-1)$. Remember that Euler's ϕ-function computes $\mathrm{ord}(\mathbb{Z}/(n))^*$. Thus $\deg \Phi_n = \phi(n)$. Further particular examples are $\Phi_4 = (x-\mathsf{i})(x+\mathsf{i}) = x^2+1 \in \mathbb{Z}[1]$ and $\Phi_6 = (x-\omega_6)(x-\bar{\omega}_6) = x^2-x+1 \in \mathbb{Z}[x]$.

Proposition 3.8.3. $x^n - 1 = \prod_{d|n} \Phi_d$

Proof. The n-th roots of unity each have order $d|n$. But a root of order d is a primitive root of unity. □

Lemma 3.8.4. *The Φ_n are monic and have integer coefficients.*

Proof. Monic follows from the representation $\Phi_n = \prod_{k\in(\mathbb{Z}/(n))^*}(x-\omega_n^k)$. The rest follows by induction over n together with long polynomial division of x^n-1 by the monic polynomial $\prod_{d|n,d<n} \Phi_d \in \mathbb{Z}[x]$. □

Proposition 3.8.5. *For all $n \in \mathbb{N}$, $\Phi_n \in \mathbb{Q}[x]$ is irreducible.*

Proof. Assume $\Phi_n \in \mathbb{Z}[x]$ were not irreducible over $\mathbb{Q}$. Then it factors over $\mathbb{Z}$, say as $\Phi_n = qr$ with $q, r \in \mathbb{Z}[x]$ each of degree at least 1. W.l.o.g. we may also assume that q is irreducible. Since Φ_n is monic the leading coefficients of q and r are ± 1 both the same, and we may thus assume that both are monic. They have complex roots the ω_n^k with $k \in (\mathbb{Z}/(n))^*$ distributed over q and r. Choose $1 \le k < n$ as small as possible such that $\epsilon := \omega^{k_q}$ with $\omega := \omega_n$ is a root of q and $\zeta = \epsilon^k$ is a root of r. We have $k > 1$, because otherwise ϵ were a multiple root of Φ_n.

Let p be a prime divisor of k. Then p does not divide n, because ϵ^k is primitive, and $\Phi_n(\omega^k) = 0$. If $q(\epsilon^p) = 0$, then $\epsilon^k = (\epsilon^p)^{k/p}$ contradicts the minimality of k. Therefore $r(\omega^p) = 0$. But $k \ge p$ is as small as possible, so $k = p$. Moreover q divides $r(x^p) \in \mathbb{Q}[x]$, because $q = \mathrm{Irr}(\epsilon/\mathbb{Q})$ and $r(\omega^p) = 0$. I.e. $r(x^p) = qs$ for some $s \in \mathbb{Q}[x]$. Since q is monic, polynomial division in $\mathbb{Z}[x]$ yields $s \in \mathbb{Z}[x]$, i.e. q divides $r(x^p)$ in $\mathbb{Z}[x]$.

Localization w.r.t. $(p) \triangleleft \mathbb{Z}$ yields $\bar{r}(x^p) = \bar{q}\bar{s}$ with $\bar{r} = x^d + \overline{r_{d-1}}x^{d-1} + \cdots + \overline{r_0}$. But in $\mathbb{F}_p = \mathbb{Z}/(p)$ we have $\bar{a}^p = \bar{a}$ for all $\bar{a} \in \mathbb{F}_p$, so

$$\bar{r}^p = x^{pd} + \bar{r}_{d-1}x^{p(d-1)} + \cdots + \bar{r}_0 = \bar{r}(x^p).$$

Therefore $\bar{r}^p$ is divisible by $\bar{q}$. But then also $\bar{r}$ and $\bar{q}$ must have a common divisor $\bar{t} \in \mathbb{F}_p[x]$ of degree at least 1. This would imply that $\bar{\Phi}_n$ which is a divisor of $x^n - \bar{1}$ has a multiple root, but the latter one is separable over $\mathbb{F}_p$, because its derivative $\partial(x^n-\bar{1}) = \bar{n}x^{n-1} \not\equiv 0$ where n is not divisible by p. This is a contradiction, so Φ_n must be irreducible. □

Proposition 3.8.6. *The Galois group of* Φ_n *is* $(\mathbb{Z}/(n))^*$.

Proof. We know that Φ_n is irreducible over $\mathbb{Q}$. Its roots are ω_n^k where $\omega_n = \mathrm{e}^{2\pi\mathrm{i}/n}$ and $k \in (\mathbb{Z}/(n))^*$. Since ω_n is primitive, an automorphism is uniquely determined by the image of ω_p. Moreover all these roots are primitive, so $\mathrm{Gal}(q/F) \cong (\mathbb{Z}/(n))^*$. □

Definition 3.8.7. *Given* $a \in F$ *where* F *is of characteristic 0, we call* $\xi \in \bar{F} \supset F$ *an* nth root (方根) of a *if it is a root of* $x^n - a \in F[x]$. *An algebraic field extension* $F \subset E$ *is a* root extension *if it is a splitting field of some* $x^n - a \in F[x]$.

An algebraic element α *over a field* F *is a* radical expression (根式) *if it is element of a tower of root extensions. The set of all radical expressions over* F *is denoted* $\operatorname{rad} F$. *We say that* $F \subset E$ *is a* radical extension (方根扩张) *iff every element* $\alpha \in E$ *is a radical expression over* F.

Remark 3.8.8. In positive characteristic we have to permit in addition roots of $x^p - x - b \in F[x]$.

Example 3.8.9. An example of a radical expression is $\sqrt[3]{1 + \sqrt{1/2}}$.

Lemma 3.8.10. *The ring of radical expressions over* F *forms a field* $\operatorname{rad} F$. □

Proposition 3.8.11 (Tower properties)**.** *Given a tower of finite algebraic extensions* $F \subset K \subset E$, *then* E/F *is a radical extension iff* K/F *and* E/K *are radical extensions.*

If $K \subset E$ *is a radical extension and* $E, F \subset L$, *then* EF/KF *is a radical extension.*

The proof is left as an exercise.

Definition 3.8.12. *A Galois extension is called* cyclic *if its Galois group is cyclic.*

Example 3.8.13. Consider again the cyclotomic polynomial $p_5(x) = x^4 + x^3 + x^2 + x + 1 \in \mathbb{Q}[x]$. We already know that its Galois group is cyclic and transitive on the roots. Therefore $\mathrm{Gal}(p_5/\mathbb{Q}) \cong C_4$.

Remark. Remember that a finite group is solvable iff (any of) its composition series has only cyclic factors of prime order.

Lemma 3.8.14. *Given a polynomial* $p := x^n - a \in F[x]$, *then its Galois group is solvable.*

Proof. Obviously the splitting field is $E_p = F(\omega_n, \xi)$ where ξ is any n-th root of a in the algebraic closure $\bar{F}$. We know that $\langle \omega_n \rangle \subset \bar{F}^*$ is cyclic and thus its Galois group $\mathrm{Gal}(\omega_n/F)$ is a factor of $(\mathbb{Z}/(n))^*$ and thus abelian. Consider now $K := F(\omega_n)$ and the extension $E = K(\xi) : K$. Its Galois group consists of permuting the roots of the minimal polynomial $q \in K[x]$. Since $p(\xi) = 0$, we know that $q|p$. But then the solutions are a subset of $\{\xi\omega_n^k : k = 0, \ldots, n-1\}$. Therefore its Galois group is also abelian and thus solvable. □

Theorem 3.8.15 (Galois)**.** *Given a separable polynomial $p \in F[x]$. Then all roots of p are radical expressions over F iff the Galois group of its splitting field is solvable.*

Proof. We will restrict to the case of characteristic 0. Let first $G := \mathrm{Gal}(p/F)$ be solvable. For every irreducible factor of p we can adjoin to F all the dth roots of unity where d is the degree of the irreducible factor. This way the Galois group of p over the extension is a factor of the original Galois group, but still solvable, i.e. there is a composition series $\{1\} = G_0 \triangleleft G_1 \triangleleft \ldots \triangleleft G_n = G$ such that each quotient G_{k+1}/G_k is cyclic. But the normal series corresponds to a tower of normal extensions $E_p := K_0 \supset K_1 \supset \cdots \supset K_n = F$ and $\mathrm{Gal}(K_k : K_{k+1}) \cong G_{k+1}/G_k$ thus cyclic.

Lemma 3.8.16 (Hilbert[16])**.** *Let $F \subset E$ be a cyclic extension and τ a generator of $\mathrm{Gal}(E : F)$, then for every $\alpha \in E$ we have the following two properties*

1. *$N(\alpha) = 1$ iff $\alpha = \tau(\gamma)/\gamma$ for some $\gamma \in E^*$,*
2. *$\mathrm{tr}(\alpha) = 0$ iff $\alpha = \tau(\gamma) - \gamma$ for some $\gamma \in E$.*

The functions N and tr are called the norm and trace, respectively, of the extension $F \subset E$. The proof of this lemma will thus be deferred to the next Section about norm and trace.

Lemma 3.8.17. *If $\mathrm{Gal}(E : F)$ is cyclic of degree n and F contains all n-th roots of unity, then $E = F(\alpha)$ for some $\alpha \in E$ with $\alpha^n \in F$.*

...

Let conversely $E := E_p = F(\xi_1, \ldots, \xi_n)$ be the splitting field of p. We can write it in the form $E_0 := F(\omega_N)$ where ω_N is a primitive N-th root of unity with N the product of all root indices in all roots of every ξ_k, and $E_{k+1} := E_k(\xi_{k+1})$. Since every ξ_k is a radical expression over F, it is also a radical expression over E_{k-1} and therefore contained in a finite tower of root extensions of E_{k-1}. Since each root extension is normal and has a cyclic subgroup, we have a normal series of

[16] David Hilbert *1/1862 in Kaliningrad (then Prussia), †2/1943

$\mathrm{Gal}(E_n : E_{n-1})$ with only abelian factors. (Note that $E_n \supset E_{n-1}$ is indeed normal, because E_0 contains the required root of unity.) By gluing together the normal series, we see that $\mathrm{Gal}(E_n : F)$ is solvable. □

Example 3.8.18.

1. Consider the polynomial $x^3 + 2 \in \mathbb{Q}[x]$. We know that its splitting field is $E := \mathbb{Q}(\sqrt[3]{2}, \omega_3)$ and the Galois group has thus 6 elements $\operatorname{ord} \mathrm{Gal}(E : \mathbb{Q}) = \dim[E : \mathbb{Q}] = 6$. On the other hand the Galois group is a subgroup of S_3 and transitive on the three roots, because the polynomial is irreducible. Therefore the Galois group is S_3 which is solvable. This coincides with the observation $\xi_1 = \sqrt[3]{2}$ and $\omega_3 = (-1 + \sqrt{-3})/2$ which implies that all three solutions are radical expressions over $\mathbb{Q}$.

2. In order to construct a polynomial whose roots cannot be computed by radicals, we need a polynomial with a Galois group that is not solvable, e.g. A_5, S_5, or bigger. We thus need an irreducible polynomial of degree 5 and need to show that its automorphisms generate A_5 or S_5. An example for which we can easily prove these properties is $Q := x^5 - 5x + 2 \in \mathbb{Q}[x]$. First observe that $Q(x - 2) = x^5 - 10x^4 + 40x^3 - 80x^2 + 75x - 20$ is irreducible by Eisenstein criterion with $p = 5$. Therefore $\mathrm{Gal}(Q/\mathbb{Q}) \subset S_5$ is transitive. Considering the polynomial as real function, we can easily see ($Q' = 5x^4 - 5$ has two roots) that it only has 3 real roots, thus 2 conjugate complex roots. But then the Galois group of Q over $\mathbb{R}$ is non-trivial and thus the Galois group over $\mathbb{Q}$ contains a transposition (complex conjugation restricted to the splitting field). Therefore it must be S_5.

Remark 3.8.19. Remember that a division ring is a unital associative (non-necessarily commutative) ring in which every nonzero element has an inverse. Examples are beside fields, also the skew-fields, e.g. $\mathbb{H}$ the quaternions.

Proposition 3.8.20 (Wedderburn)**.** *Every finite division ring is a field.*

Proof. Let D be a finite division ring and $F := \mathrm{cent}(D) := \{z \in D : \forall x \in D : zx = xz\}$ be its center. Clearly $F \subset D$ is a subfield and thus D a vector space over F. Since D is finite $n := \dim_F D < \infty$. We want to show that $n = 1$.

Let $|F| = q$, so that $|D| = q^n$. The center of $D^* = D \setminus \{0\}$ has the $q - 1$ elements F^*. The centralizer $Z := Z_D(a) := \{x \in D : ax = xa\}$ of $a \in D$ in $D \setminus \{0\}$ is $Z \setminus \{0\}$. Note that $Z \subset D$ is a subring, a division ring, and contains F. Hence $d := \dim_F Z$ divides n and in particular $|Z| = q^d$. So the centralizer of $a \in D^*$ has $q^d - 1$ elements and the conjugacy class $(q^n - 1)/(q^d - 1)$ elements.

Moreover $q^d < q^n$ if $a \in D \setminus F$. Hence the class equation of the multiplicative group D^* reads

$$q^n - 1 = |D^*| = (q-1) + \sum_{|C|>1} \frac{q^n - 1}{q^d - 1}$$

where the sum runs over the non-trivial conjugacy classes and $d|n$ depends on the class. Now $q^n - 1 = \prod_{d|n} \Phi_d(q)$ where $\Phi_d \in F[x]$ is the cyclotomic polynomial and d runs over the positive divisors of n that are smaller than n. But for $d < n$ and $d|n$, we can further decompose as

$$\begin{aligned} q^n - 1 &= \Phi_n(q) \prod_{c|n, c<n} \Phi_c(q) \\ &= \Phi_n(q) \prod_{c|d} \Phi_c(q) \prod_{c|n, c<n, c\not|d} \Phi_c(q) \\ &= \Phi_n(q)(q^d - 1) \prod_{c|n, c<n, c\not|d} \Phi_c(q) \end{aligned}$$

where $\Phi_n(q)$ divides $q^n - 1$ and $(q^n - 1)/(q^d - 1)$. Therefore $\Phi_n(q)$ divides $q - 1$. But on the other hand $\Phi_n(q) > q - 1$ for $n > 1$. Therefore $n = 1$ which completes the proof. □

It is also possible to further exploit the cyclotomic polynomials and obtain the following two results from number theory.

Theorem 3.8.21 (Dirichlet[17])**.** *Every arithmetic progression $a_n = an + b$ where $a, b \in \mathbb{N}$ and $\gcd(a, b) = 1$ contains infinitely many primes.* □

Proposition 3.8.22. *Every finite abelian group occurs as Galois group of a finite extension of $\mathbb{Q}$.*

Idea of proof. Let $A = C_{n_1} \oplus \cdots \oplus C_{n_k}$ be a finite abelian group and $n_i > 1$. By Dirichlet's theorem there are distinct primes $p_i \in \mathbb{P}$ such that $p_i \equiv 1 \pmod{n_1 n_2 \cdots n_k}$ for all $1 \leq i \leq k$. Let now $N := p_1 p_2 \cdots p_k$ be their product. We consider the extension $E := \mathbb{Q}(\omega_N)$ by a primitive N-th root of unity.

By Proposition 3.8.6, $\mathrm{Gal}(E : \mathbb{Q}) \cong (\mathbb{Z}/(N))^*$. Remember that for k, l relatively prime $\mathbb{Z}/(k) \times \mathbb{Z}/(l) \cong \mathbb{Z}/(kl)$ and $(u, v) \in \mathbb{Z}/(k) \times \mathbb{Z}/(l)$ is a unit iff u is relatively prime to k and v is relatively prime to l, i.e. $(\mathbb{Z}/(kl))^* \cong (\mathbb{Z}/(k))^* \times (\mathbb{Z}/(l))^*$. Therefore $\mathrm{Gal}(E : \mathbb{Q}) = (\mathbb{Z}/(n))^* \cong (\mathbb{Z}/(p_1))^* \oplus \cdots \oplus (\mathbb{Z}/(p_k))^*$. But the latter are $(\mathbb{Z}/(p))^* \cong C_{p-1}$, because $\mathbb{F}_p := \mathbb{Z}/(p)$ is a field. But by construction $n_i|(p_i - 1)$, hence $(\mathbb{Z}/(p_i))^*$ has a subgroup H_i of index n_i. This means $(\mathbb{Z}/(p_i))^*/H_i \cong C_{n_i}$. Moreover $\mathrm{Gal}(E : \mathbb{Q}) \cong (\mathbb{Z}/(p_1))^* \oplus \cdots \oplus (\mathbb{Z}/(p_k))^*$ has a subgroup $H_1 \oplus \cdots \oplus H_k$ with

[17] Peter G.L. Dirichlet *1805/2 Düren/France, †1859/5

$\mathrm{Gal}(E:\mathbb{Q})/(H_1 \oplus \dots H_k) \cong A$. But the fixed field $K \subset E$ of $H := H_1 \oplus \cdots \oplus H_k$ is a finite Galois extension of $\mathbb{Q}$ and thus $\mathrm{Gal}(K:\mathbb{Q}) \cong A$ as required. □

3.8.99 Exercises

Exercise 3.8.1. a. Find $\Phi_n \in \mathbb{Q}[x]$ for all $n \leq 10$,

b. Find Φ_{12} and $\Phi_{18} \in \mathbb{Q}[x]$.

Exercise 3.8.2. a. Show that $\Phi_n(0) = \pm 1$,

b. show that $\Phi_{2k+1}(0) = 1$ if $k \geq 1$.

Remark 3.8.23 (Warning)**.** Not for every Φ_n are all the coefficients $0, \pm 1$. Unfortunately to find a counter example requires to search among the irreducible factors of $x^n - 1 \in \mathbb{Q}[x]$.

Exercise 3.8.3. Let $p^2|n$ for some prime $p \in \mathbb{P}$. Show that the sum of all complex primitive n-th roots of unity is 0.

Exercise 3.8.4. a. Show that $\mathbb{Q}(\omega_m)\mathbb{Q}(\omega_n) = \mathbb{Q}(\omega_l)$ where $l = \mathrm{lcm}(m,n)$ is the least common multiple.

b. Show that $\mathbb{Q}(\omega_m) \cap \mathbb{Q}(\omega_n) = \mathbb{Q}(\omega_d)$ where $d := \gcd(m,n)$ is the greatest common divisor.

Exercise 3.8.5. Find the smallest $n \in \mathbb{N}$ such that $\mathrm{Gal}(\mathbb{Q}(\omega_n):\mathbb{Q})$ is not cyclic.

Exercise* 3.8.6. Prove that every finitely generated module[18] over any division ring has a finite basis and that all bases have the same cardinality (number of elements).

Exercise 3.8.7. Using the results of the previous Exercise 3.8.6 show that for a tower of division rings $D \subset K \subset E$, $\dim_D E = (\dim_D K)(\dim_K E)$ and in particular the left side is infinite iff at least one of the two factors on the right side is infinite.

Exercise 3.8.8. Show the tower properties of radical extensions, i.e.

a. Given a tower of finite algebraic extensions $F \subset K \subset E$, then E/F is a radical extension iff K/F and E/K are radical extensions.

b. If $K \subset E$ is a radical extension and $E, F \subset L$, then EF/KF is a radical extension.

[18] the analogue of a vector space

3.9 Norm (赋范) and trace (迹)

Definition 3.9.1. *Given a ring R. A* norm/valuation *is a map $n\colon R \to (-\infty, \infty)$ with the property that $n(\alpha) = 0$ iff $\alpha = 0$, and $n(\alpha\beta) = n(\alpha)n(\beta)$.*

Given an algebra A over a field F. A norm is a map $n\colon A \to F$ with the property $n(\alpha) = 0$ iff $\alpha = 0$, and $n(\alpha\beta) = n(\alpha)n(\beta)$.

Example 3.9.2. Given a finite dimensional field extension $F \subset E$. We consider the F-linear map $M_\alpha\colon E \to E : v \mapsto \alpha v$ of multiplication with $\alpha \in E$. The determinant $N\colon E \to F : \alpha \mapsto \det M_\alpha$ of M_α is an element in F and moreover N a norm, because $\det M_{\alpha\beta} = \det(M_\alpha M_\beta) = (\det M_\alpha)(\det M_\beta)$ and $\det M_\alpha = 0$ implies that α is a zero-divisor, hence 0.

The operation $\alpha \mapsto \operatorname{tr} M_\alpha$ is a *trace*.

Remark 3.9.3. Note that the existence of a norm implies that the ring is a domain. The normed algebras are also called *division algebras*, because every non-zero element has an inverse if there is an identity in the algebra.

Example 3.9.4. The *real division algebras* are $\mathbb{R}$, $\mathbb{C} = \mathbb{R}(\mathsf{i})$, $\mathbb{H} = \mathbb{R}(\mathsf{i},\mathsf{j})$, and $\mathbb{O} = \mathbb{R}(\mathsf{i},\mathsf{j},\mathsf{E})$ (Octonions, discovered by Graves[19] and independently by Cayley[20]). While $\mathbb{R}$ is linear ordered, $\mathbb{C}$ is algebraically closed, $\mathbb{H}$ is associative but not commutative, and $\mathbb{O}$ is not even associative.

The question may occur what is the relation of norm/ trace to the Galois/ automorphism group of the extension. This is answered with the following two statements:

Lemma 3.9.5. *If $F \subset E$ is a finite extension of degree n, $\alpha \in E$ and $q = \operatorname{Irr}_F(\alpha)$ the monic minimal polynomial of α of degree d, then*

$$\det(x\mathbb{1} - M_\alpha) = q^{n/d}$$

where in particular d divides n.

Proof. Remember that for every $a \in F$, $M_{a\alpha} = aM_\alpha$, as well as for $\beta \in E$, $M_{\alpha+\beta} = M_\alpha + M_\beta$. Hence $f(M_\alpha) = M_{f(\alpha)}$ for every $f \in F[x]$. In particular $q(M_\alpha) = M_{q(\alpha)} = 0$, i.e. q is the minimal polynomial of M_α. As opposed to arbitrary minimal polynomials, we know in addition that q is irreducible. In the factorization of ch_α into irreducible polynomials over F, we see that each factor is divisible by the minimal polynomial q of one of its roots and thus $\mathrm{ch}_\alpha = q^{n/d}$. □

[19] John T. Graves *1806/12 Dublin/Ireland, †1870/3

[20] Arthur Cayley *1821/8 in London/GB, †1895/1

Proposition 3.9.6. *Let $F \subset E$ be a finite extension of degree n and $\alpha_1, \dots, \alpha_s \in \bar{F}$ be the distinct conjugates of $\alpha \in E$. Let further $\phi_1, \dots, \phi_t$ be the distinct F-homomorphisms of E into $\bar{F}$. Then s and t divide n as well as*

$$N(\alpha) = (\alpha_1 \dots \alpha_s)^{n/s} = (\phi_1(\alpha) \dots \phi_t(\alpha))^{n/t},$$
$$\operatorname{tr}(\alpha) = \tfrac{n}{s}(\alpha_1 + \dots + \alpha_s) = \tfrac{n}{t}(\phi_1(\alpha) + \dots + \phi_t(\alpha)).$$

Some books use these properties as definition of norm and trace. The linearity of the trace as well as the multiplicativity of the norm follow then from the homomorphism properties of the maps ϕ_k.

Proof. The conjugates α_k of $\alpha \in E$ are the roots of its minimal polynomial $q = \operatorname{Irr}(\alpha/F)$ which all have the same multiplicity m, i.e.

$$q = (x - \alpha_1)^m \cdots (x - \alpha_s)^m = x^{ms} - m(\alpha_1 + \dots + \alpha_s) + \dots + (-1)^{ms}(\alpha_1 \dots \alpha_s)^m.$$

But then $ms = [F(\alpha) : F]$ divides $n = [E : F]$ with $l := [E : F(\alpha)] = \frac{n}{ms}$. So by the previous lemma $\det(x\mathbb{1} - M_A) = q^l$ with the absolute term

$$(-1)^n N(\alpha) = (-1)^{lms}(\alpha_1 \dots \alpha_s)^{ml}.$$

The linear term is

$$-\operatorname{tr}(\alpha) = -lm(\alpha_1 + \dots + \alpha_s).$$

Finally let $t := [E : F]_s$ the separability degree of $F \subset E$. We know that $t|n$ as well as $k := [E : F(\alpha)]_s = \frac{t}{s}$. But then

$$\phi_1(\alpha) \dots \phi_t(\alpha) = (\alpha_1 \dots \alpha_s)^k$$

as well as

$$\phi_1(\alpha) + \dots + \phi_t(\alpha) = k(\alpha_1 + \dots + \alpha_s).$$

This completes the proof. □

Corollary 3.9.7. *Let $F \subset E$ be a finite extension of degree n and $\alpha \in E$. Then N and* tr *simplify as follows:*

1. *For $a \in F$, then $N(a) = a^n$, $\operatorname{tr}(a) = na$.*
2. *If $E = F(\alpha)$ is separable, then $N(\alpha) = \alpha_1 \dots \alpha_n$ and $\operatorname{tr}(\alpha) = \alpha_1 + \dots + \alpha_n$ where α_k are the conjugates of α.*
3. *If E is not separable, then $\operatorname{tr} \equiv 0$.*

4. If $F \subset E$ is Galois with Galois group G, then

$$N(\alpha) = \prod_{\sigma \in G} \sigma(\alpha),$$
$$\mathrm{tr}(\alpha) = \sum_{\sigma \in G} \sigma(\alpha).$$

The proof is left as an exercise.

As shown in the last proposition, norm and trace depend on the separability degree of the finite extension $F \subset E$. So it is not surprising that it fulfills similar tower properties as the separability degree:

Proposition 3.9.8 (Tower property)**.** *Let $F \subset K \subset E$ be finite extensions. Then*

$$N_F^E(\alpha) = N_F^K(N_K^E(\alpha)),$$
$$\mathrm{tr}_F^E(\alpha) = \mathrm{tr}_F^K(\mathrm{tr}_K^E(\alpha))$$

for all $\alpha \in E$.

Also this proof is left as an exercise.

Remember that we have some properties of cyclic extensions missing for the proof of Galois' Theorem (about solvability of polynomial equations with radical expressions). In order to prove Hilbert's 90th Theorem (Lemma 3.8.16/3.9.10), we first need to show the following property:

Lemma 3.9.9. *Let $F \subset E, K$ be field extensions. The distinct F-homomorphisms of K into E are linearly independent over E.*

Proof. Assume there is an equality $c_1\phi_1 + \cdots + c_n\phi_n = 0$ in which not all the $c_k \in E$ vanish and $\phi_1, \ldots, \phi_n$ are distinct F-homomorphisms of K into E. Among those relations there is one in which n is smallest, i.e. all $c_k \neq 0$ and $n \geq 2$. Then

$$c_1\phi_1(\alpha)\phi_1(\beta) + \cdots + c_n\phi_n(\alpha)\phi_n(\beta) = (c_1\phi_1 + \cdots + c_n\phi_n)(\alpha\beta) = 0,$$
$$c_1\phi_n(\alpha)\phi_1(\beta) + \cdots + c_n\phi_n(\alpha)\phi_n(\beta) = \phi_n(\alpha)(c_1\phi_1 + \cdots + c_n\phi_n)(\beta) = 0$$

for all $\alpha, \beta \in K$. Their difference is

$$c_1(\phi_1 - \phi_n)(\alpha)\phi_1(\beta) + \cdots + c_n(\phi_{n-1} - \phi_n)(\alpha)\phi_{n-1}(\beta) = 0.$$

Since this is a function in $\beta \in K$, we also obtain

$$c_1(\phi_1 - \phi_n)(\alpha)\phi_1 + \cdots + c_n(\phi_{n-1} - \phi_n)(\alpha)\phi_{n-1} = 0.$$

But this means that we have found a (nontrivial) algebraic relation with less homomorphisms, which is a contradiction. □

Lemma 3.9.10 (Hilbert[16]). *Let $F \subset E$ be a cyclic extension and τ a generator of* $\mathrm{Gal}(E:F)$*, then for every $\alpha \in E$ we have the following two properties*

1. $N(\alpha) = 1$ *iff* $\alpha = \tau(\gamma)/\gamma$ *for some* $\gamma \in E^*$,
2. $\mathrm{tr}(\alpha) = 0$ *iff* $\alpha = \tau(\gamma) - \gamma$ *for some* $\gamma \in E$.

Proof. Let G be the Galois group of $E:F$. If $\gamma \in E^*$, then

$$N(\tau\gamma) = \prod_{\sigma\in G} \sigma\circ\tau(\gamma) = \prod_{\sigma\in G}\sigma(\gamma) = N(\gamma).$$

Hence for $\gamma \neq 0$, we have $N(\tau\gamma/\gamma) = 1$.

Conversely assume that $N(\alpha) = 1$. By the previous lemma we know that $\mathrm{Id}, \tau, \tau^2, \ldots, \tau^{n-1}$ are E-linearly independent when $n = [E:F]$. But then

$$\phi := \mathrm{Id} + \alpha\tau + \alpha\tau(\alpha)\tau^2 + \cdots + \alpha\tau(\alpha)\ldots\tau^{n-2}(\alpha)\tau^{n-1} \neq 0$$

as well as

$$\delta := \phi(\beta) \neq 0$$

for some $\beta \in E$. Since

$$N(\alpha) = \alpha\tau(\alpha)\ldots\tau^{n-1}(\alpha),$$
$$\alpha\tau(\delta) = \alpha\tau(\beta) + \alpha\tau(\alpha)\tau^2(\beta) + \cdots + \alpha\tau(\alpha)\ldots\tau^{n-1}(\alpha)\beta = \delta$$

Hence $\alpha = \tau(\gamma)/\gamma$ for $\gamma = \delta^{-1}$.

Similarly for $\gamma \in E$ we obtain

$$\mathrm{tr}(\tau\gamma) = \sum_{\sigma\in G}\sigma\circ\tau(\gamma) = \sum_{\sigma\in G}\sigma(\gamma) = \mathrm{tr}(\gamma)$$

and so $\mathrm{tr}(\tau\gamma - \gamma) = 0$.

Conversely assume that $\mathrm{tr}(\alpha) = 0$. Again $\mathrm{Id}, \tau, \ldots, \tau^{n-1}$ are E-linearly independent and so

$$\mathrm{tr} = \mathrm{Id} + \tau + \cdots + \tau^{n-1} \neq 0$$

as well as

$$\mathrm{tr}(\beta) \neq 0$$

for some $\beta \in E$. Let now

$$\delta := \alpha\tau(\beta) + (\alpha + \tau(\alpha))\tau^2(\beta) + \cdots + (\alpha + \tau(\alpha) + \cdots + \tau^{n-2}(\alpha))\tau^{n-1}(\beta)$$

and observe that

$$\tau(\delta) = \tau(\alpha)\tau^2(\beta) + (\tau(\alpha) + \tau^2(\alpha))\tau^3(\beta) + \cdots - \alpha\beta$$

where in the last term we used $\operatorname{tr}(\alpha) = 0$ as well as $\tau^n = \mathrm{Id}$. Hence

$$\delta - \tau(\delta) = \alpha\tau(\beta) + \alpha\tau^2(\beta) + \cdots + \alpha\tau^{n-1}(\beta) = \alpha \operatorname{tr}(\beta)$$

and so $\alpha = \tau(\gamma) - \gamma$ for $\gamma = -\delta/\operatorname{tr}(\beta)$. □

Remark 3.9.11. If you wish to extend the Galois theorem about solvability of polynomial equations by radical expressions to fields not of characteristic 0, then you need a Property corresponding to Lemma 3.8.17. It turns out that the original statement holds as long as the extension degree $n = [E : F]$ is not divisible by the characteristic. If it is, the corresponding property is the Artin–Schreier Theorem and it also replaces the standard polynomial for root extensions by $x^n - x - b \in F[x]$.

3.9.99 Exercises

Exercise 3.9.1. Find N for $E := \mathbb{Q}(\alpha) \subset \mathbb{C}$ with

a. $\alpha = \sqrt{n}$ for some $n \in \mathbb{N}^*$,

b. $\alpha = \mathsf{i}\sqrt{n}$ for some $n \in \mathbb{N}^*$,

c. $\alpha = \sqrt{2} + \sqrt{3}$,

d. $\alpha = \sqrt{2} + \mathsf{i}\sqrt{3}$.

Exercise 3.9.2. Define the unit octonions as $\mathbb{O}_1 := \langle -1, \mathsf{i}, \mathsf{j}, \mathsf{E} : \mathsf{i}^2 = -1 = \mathsf{j}^2 = \mathsf{E}^2, (-1)^2 = \mathrm{id}, \ldots \rangle_{\mathbb{O}}$ together with the multiplication law

id	i	j	k	E	I	J	K
i	−1	k	−j	I	−E	−K	J
j	−k	−1	i	J	K	−E	−I
k	j	−i	−1	K	−J	I	−E
E	−I	−J	−K	−1	i	j	k
I	E	−K	J	−i	−1	−k	j
J	K	E	−I	−j	k	−1	−i
K	−J	I	E	−k	−j	i	−1

a. Check that $\langle u, v \rangle_{\mathbb{O}}$ for any $u, v \in \mathbb{O}_1$ forms a multiplicative set isomorphic to a subgroup of the unit quaternions and thus is a group (including associativity),

b. Note that $\mathbb{O}_1$ is not associative (thus does not form a group). Check that instead it fulfills the alternative laws

$$(uv)v = u(vv),$$
$$(uu)v = u(uv)$$

for all $u, v \in \mathbb{O}_1$.

c. Define $\mathbb{O} := \mathbb{R}[\mathbb{O}_1]/(-\mathrm{id} = -1)$ where $\mathbb{O}$ is an $\mathbb{R}$-algebra and conclude that it fulfills the same alternative laws. Note that $\langle u, v\rangle_{\mathbb{O}}$ for $u, v \in \mathbb{O}$ generates a subalgebra isomorphic to a subalgebra of the quaternions (thus being associative).

d. Define the norm of an octonion as $|z|^2 := z\bar{z}$ for $z \in \mathbb{O}$ and $\overline{\pm 1} = \pm 1$, $\bar{\imath} = -\mathsf{i}$, $\bar{\jmath} = -\mathsf{j}$, $\bar{\mathsf{E}} = -\mathsf{E}$ and correspondingly for the other units. Verify that $|zw| = |z|\,|w|$ as well as $|z| = 0$ iff $z = 0$.

e. Conclude that $\mathbb{O}$ is a division algebra. (What are the inverse elements?)

Exercise* 3.9.3. Show the simplified formulas for trace and norm in Corollary 3.9.7, i.e. let $F \subset E$ be an extension of finite degree.

a. For $a \in F$, then $N(a) = a^n$, $\mathrm{tr}(a) = na$.

b. If $E = F(\alpha)$ is separable, then $N(\alpha) = \alpha_1 \dots \alpha_n$ and $\mathrm{tr}(\alpha) = \alpha_1 + \dots + \alpha_n$ where α_k are the conjugates of α.

c. $\mathrm{tr} \equiv 0$ iff E is not separable over F.

d. If $F \subset E$ is Galois with Galois group G, then

$$N(\alpha) = \prod_{\sigma \in G} \sigma(\alpha),$$
$$\mathrm{tr}(\alpha) = \sum_{\sigma \in G} \sigma(\alpha).$$

Exercise 3.9.4. Show the tower properties for norm and trace, i.e. let $F \subset K \subset E$ be finite extensions, then

$$N_F^E(\alpha) = N_F^K(N_K^E(\alpha)),$$
$$\mathrm{tr}_F^E(\alpha) = \mathrm{tr}_F^K(\mathrm{tr}_K^E(\alpha))$$

for every $\alpha \in E$.

3.10 Geometric constructions (尺规作图)

Theorem 3.10.1. *Given a unit length in the plane, then the points that can be constructed (non-stochastically with finitely many steps each) with ruler and compass are those with coordinates in* $\mathrm{rad}_2\,\mathbb{Q} \supset \mathbb{Q}$ *the algebraic numbers that are radicals with square roots only.*

Idea of proof. The construction of solutions of linear and quadratic equations is left as an exercise.

Conversely, consider a construction with a ruler. We can construct the intersection of two straight lines which has coordinates the solution of two linear equations. Given in addition a compass, we can construct intersections of two circles or of a circle and a straight line. Both have coordinates the solutions of quadratic or biquadratic equations. □

Example 3.10.2 (Impossibility of the classical Greek construction problems)**.**

1. Consider the question of *doubling the cube*. Given the side length of a cube, we are looking for the side length of a cube of doubled volume. W.l.o.g. the given cube has side length 1, but then we need to construct the number $\sqrt[3]{2} \in \mathbb{Q}(\sqrt[3]{2})$. But the latter has degree 3 over the rationals and is thus not in $\mathrm{rad}_2\,\mathbb{Q}$.

2. Consider next the *trisection of an arbitrary angle* α. For particular values, e.g. $\alpha = 2\pi$, this is possible, because all we have to do is construct an angle of $2\pi/3 = 120^o$. On the other hand for $\alpha = \pi/3$, we need to construct the sine and cosine of $\alpha/3 = \pi/9$. However, the cosine is root of the cubic polynomial $4x^3 - 3x - \frac{1}{2} \in \mathbb{Q}[x]$. Since this polynomial has no root in $\mathbb{Q}$ it is irreducible, each root has degree 3 over $\mathbb{Q}$, and is thus not an element of $\mathrm{rad}_2\,\mathbb{Q}$.

3. Consider finally the *quadrature of the circle* when the radius is given. W.l.o.g. we can assume that the radius is 1. But then we would need to construct the number $\sqrt{\pi}$ which is not even algebraic over $\mathbb{Q}$.[21]

Another classical question is that of constructability of regular n-gons (given the circumscribed circle). From high-school mathematics you probably remember how to construct a regular triangle, square, or hexagon. In an advanced course you may have seen the construction of a regular 5-gon, but why is there no algorithm for a regular 9-gon?

[21] Lindemann–Weierstrass theorem

Proposition 3.10.3 (Gauss[22], Wantzel[23]). *A regular n-gon is constructible with ruler and compass iff $n = 2^k p_1 \dots p_m$ where $p_i = 2^{2^{q_i}} + 1 \in \mathbb{P}$ with $q_i \in \mathbb{N}$ distinct integers.*

Remark 3.10.4. The p_is are called Fermat[3] primes and it is easy to check that $2^{2^0} + 1 = 3$, $2^{2^1} + 1 = 5$, $2^{2^2} + 1 = 17$, $2^{2^3} + 1 = 257$, and $2^{2^4} + 1 = 65537$ are primes. So far no further Fermat primes have been found, but a proof that these 5 are all Fermat primes is still missing.

Proof. A regular n-gon is constructible from its radius iff the primitive n-th root of unity $\omega_n = \mathrm{e}^{2\pi \mathsf{i}/n}$ is constructible. By Proposition 3.8.6, ω_n has degree $\phi(n)$ over $\mathbb{Q}$. Writing n as a product $n = 2^m p_1^{m_1} \dots p_k^{m_k}$ of a power of 2 and distinct odd primes, we obtain that

$$\phi(n) = 2^{m-1} p_1^{m_1-1}(p_1 - 1) \dots p_k^{m_k-1}(p_k - 1).$$

Hence $\phi(n)$ is a power of 2 iff $m_1 = m_2 = \dots = m_k = 1$ and $p_1 - 1, \dots, p_k - 1$ are powers of 2. But then the p_i must be distinct Fermat primes by the following Lemma:

Lemma 3.10.5. *If $2^k + 1$ is a prime, then $k = 2^n$.*

Proof. If k is not a power of 2, then $k = 2^n j$ where j is odd and at least 3. Then every $m^j + 1$ is divisible by $m + 1$ and thus $(2^{2^n})^j + 1$ is divisible by $2^{2^n} + 1$. □

Remark 3.10.6. It is also possible to do a relative construction theory, i.e. beside a line segment of length 1, we may also be given further line segments/ angles which all together generate the field $F \subset \mathbb{C}$. The constructible points are now these with coordinates in $\mathrm{rad}_2 F \subset \mathbb{C}$. Note that $\mathbb{Q} \subset F$ as the prime field, i.e. the absolutely constructible numbers are also relatively constructible.

3.10.99 Exercises

Exercise* 3.10.1. Find the construction of the regular 17-gon with ruler and compass.

Exercise 3.10.2. Given a unit line segment and ruler and compass, show that the following are constructible

0. rational numbers,

a. the imaginary unit i,

[22] Carl-Friedrich Gauß 4/1777 in Braunschweig (now Germany), †2/1855
[23] Pierre L. Wantzel 6/1814 in Paris/France, †5/1848

b. given a line segment of length $a > 0$, then $\sqrt{a}$ is constructible,

c. given a point $z \in \mathbb{C}$, then the point $\sqrt{z} \in \mathbb{C}$, i.e. each of the two square roots are constructible.

3.11 Algebraic integers (代数整数)*

3.12 Outlook: Algebraic geometry (代数几何)

Definition 3.12.1. *Given a field F, then an* algebraic set/variety (代数变形) *V is the joint zero locus of a family $\mathcal{P} \subset F[x_1, \ldots, x_n]$, i.e. $V(\mathcal{P}) = \{\mathbf{x} \in \bar{F}^n : \forall p \in \mathcal{P} : p(\mathbf{x}) = 0\}$.*

Example 3.12.2. 0. Consider the trivial polynomial $p = 0 \in F[x_1, \ldots, x_n]$, then $V(0) = \bar{F}^n$ i.e. the whole (affine) space. Conversely for $p = 1$ we obtain $V(1) = \emptyset$, i.e. the other trivial set.

1. Consider the field $\mathbb{R}$ of real numbers and the polynomial $p := x_1^2 + \cdots + x_n^2 - 1$. Its zero locus is $\mathbb{S}^{n-1} := V(p) \cap \mathbb{R}^n$ the $n-1$-dimensional unit sphere (球面) embedded into $\mathbb{R}^n$.[24]

2. Consider the polynomial $p := x_1 \cdots x_n \in F[x_1, \ldots, x_n]$. Its zero locus is the union of the coordinate hyper-planes (坐标超平面) through 0.

Remark 3.12.3. Note that we permit the zeros to be in the algebraic closure. This is necessary, because otherwise quite different polynomials, e.g. $x^3 - 2 \in \mathbb{Q}[x]$ and $x^2 - 6 \in \mathbb{Q}[x]$ would produce the same set $\emptyset \subset \mathbb{Q}$. The question whether an algebraic variety over $\mathbb{Q}$ has rational points (有理点) can be quite hard, e.g. Fermat's Last Theorem (费马大定理, $p := x^n + y^n - z^n \in \mathbb{Q}[x, y, z]$).

Remark 3.12.4. We can extend the family of polynomials generating the algebraic variety to an ideal, because evaluation is a ring homomorphism. Secondly, there is some advanced property of polynomial rings called Hilbert's Basis Theorem **??** (希耳伯特所产生定理) which states that every ideal in a polynomial ring (with finitely many indeterminates) over a field has a finite generating system. Therefore it is sufficient to consider varieties generated by finitely many polynomials.

Example 3.12.5 (Projective spaces (射影空间))**.** Given a field F, there is a free action of F^* on $F^{n+1} \setminus 0$ by component-wise multiplication. Therefore the quotient (商空间) is a nice space $\mathbb{P}^n F := (F^{n+1} \setminus 0)/F^*$ called projective space.

More particularly $\mathbb{P}^1\mathbb{R} = \mathbb{R} \cup \{\infty\} \approx \mathbb{S}^1$ as well as $\mathbb{P}^1\mathbb{C} = \mathbb{C} \cup \{\infty\} \approx \mathbb{S}^2$.

[24] V is its extension to $\mathbb{C}^n$, because $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$.

Definition 3.12.6. *1. A polynomial is called* homogeneous (齐次) *if the total degree of each of its monomials is the same.*

2. A projective variety (射影簇) *over a field F is the common zero locus in $\bar{F}^{n+1}$ of a family $\mathcal{P}$ of homogeneous polynomials in $F[x_0, x_1, \ldots, x_n]$ modulo F^*, i.e. $V_{\mathbb{P}}(\mathcal{P}) \subset \mathbb{P}^n\bar{F}$.*

3. A homogeneous ideal *is an ideal generated by homogeneous polynomials.*

Remark 3.12.7. The reason to consider projective varieties instead of algebraic ones is that $\mathbb{P}^nF$ is *compact* (say for $F = \mathbb{R}, \mathbb{C}, \mathbb{H}$ or finite) while F^n generally is not (in usual topology if F is infinite). Therefore the projective varieties are compact while the algebraic ones may or may not be compact. In what follows we will mainly focus on algebraic varieties, because here the proofs are a bit easier. The advanced reader will however be able to transfer most of the proofs also to projective varieties/ homogeneous ideals.

Remark 3.12.8. Note that the variety to the intersection or product of ideals $I = I_1 \cap I_2$ (or $I = I_1I_2$) is the union of the varieties $V(I) = V(I_1) \cup V(I_2)$. While the product of ideals is only defined for finitely many ideals, the intersection of arbitrarily many ideals may not correspond to the union of all their varieties.

Conversely the ideal generated by the union of ideals $I = (I_\alpha : \alpha \in A) = \sum_{\alpha\in A} I_\alpha$ produces the intersection of the varieties $V(I) = \bigcap_{\alpha\in A} V(I_\alpha)$.

Example 3.12.9. As an example of a set that is not an algebraic variety, consider the set $S = \pi\mathbb{Z} \subset \mathbb{C}^1$. While it is the zero locus of the *analytic function (解析函数)* $\sin x$, it is not the zero locus of any family of polynomials $\mathcal{P} \subset \mathbb{R}[x]$, because such a family would have to contain a non-zero polynomial with infinitely many roots (over the field $\mathbb{C}$).

Remark 3.12.10. The previous two properties of the family of algebraic varieties in $\bar{F}^n$ resembles those of closed sets and ideed, we can define the *Zariski*[25] *topology (扎里斯基拓扑)* on $\bar{F}^n$ as the topology with the algebraic varieties as closed sets. (The open sets are thus the complements of algebraic varieties.) Unfortunately this topology is not *Hausdorff*[26] *(豪斯多夫拓扑)*, namely given any two (different) points $\mathbf{x}_1, \mathbf{x}_2 \in \bar{F}^n$, then there are no algebraic varieties $\mathbf{x}_k \notin V_k \subset \bar{F}^n$ with $\mathbf{x}_{2-k} \in V_k$ and $V_1 \cup V_2 = \bar{F}^n$.[27]

[25] Oscar Zariski *1899/4 in Kobrin/Russia, †1986/7

[26] Felix Hausdorff *1868/11 in Breslau/Prussia, †1942/1

[27] It means that we are not able to find disjoint open neighborhoods $x_k \in U_k \subset \bar{F}^n, U_1 \cap U_2 = \emptyset$. This is quite opposite to usual, say metric, topologies.

Remember that an ideal $I \triangleleft R$ is called *irreducible (不可约)* iff it cannot be written as the intersection $I = I_1 \cap I_2$ of two strictly larger ideals $I \neq I_k \triangleleft R$. Correspondingly, we define an irreducible variety as follows:

Definition 3.12.11. *Given an algebraic variety $V \subset \bar{F}^n$, we call V* irreducible (不可约) *iff it cannot be written as the union of two strictly smaller algebraic varieties.*

Now we obtain the following result:

Proposition 3.12.12. *Given an ideal $I \triangleleft F[x_1, \ldots, x_n]$, then its algebraic variety $V(I)$ is irreducible iff I is irreducible.*

Proof. This follows from the correspondence intersections of ideals to unions of varieties. □

Some textbooks denote the (possibly) reducible algebraic varieties as algebraic sets while they keep the word algebraic variety to mean irreducible variety.

The next question we want to approach is how far we can reconstruct the ideal from the (embedded) algebraic variety. If $V(I)$ is the joint zero locus of the functions in I, then the following definition is at hand.

Definition 3.12.13. *Given an algebraic variety $V \subset \bar{F}^n$, we denote the* vanishing ideal (消失理想) $I(V) := \{p \in F[x_1, \ldots, x_n] : p(V) = 0\}$. *The* coordinate ring (坐标环) *of V is $R(V) := F[x_1, \ldots, x_n]/I(V)$.*

Example 3.12.14. Consider the polynomials $p = x_1^2 + x_2^2 - 1 \in \mathbb{R}[x_1, x_2]$ and p^2. Both have the same zero locus $V := V(I) \cap \mathbb{R}^2 = \mathbb{S}^1$. Thus the vanishing ideal $I(V) := \{f \in \mathbb{R}[x_1, x_2] : f|V = 0\}$ must be the same. Since the former polynomial is irreducible, we see that $I(\mathbb{S}^1) = (p) \triangleleft \mathbb{R}[x_1, x_2]$. The coordinate ring is correspondingly $R(\mathbb{S}^1) = F[x_1, x_2]/(p)$. This is a domain, because p is irreducible (and $F[x_1, \ldots, x_n]$ a UFD).

Remark 3.12.15. The coordinate ring can also be understood in the following way. We see that every element $f \in R[x_1, \ldots, x_n]$ maps the algebraic variety $V \subset \bar{F}^n$ to $f(V) \subset \bar{F}$. But apparently $f(V)$ only depends on the restriction of f to V, thus we can divide out the vanishing ideal $I(V) \triangleleft F[x_1, \ldots, x_n]$ of V. The name *coordinate* ring is now, because the standard coordinates $x_1, \ldots, x_n$ generate $F[x_1, \ldots, x_n]$ and thus their images also $R(V)$.

Note that even though Zariski topology is not Hausdorff, the coordinate ring separates points, i.e. for every two different $\mathbf{x}_i \in V$ there are two functions $f_i \in R(V)$ such that $f_i(\mathbf{x}_j) = \delta_{ij}$.

The transition from (p^2) to (p) is captured by the following process:

Proposition 3.12.16. *Given an ideal $J \triangleleft F[x_1, \ldots, x_n]$, then the vanishing ideal of the algebraic variety $V := V(J)$ is*

$$I(V) = \sqrt{J} = \{p \in F[x_1, \ldots, x_n] : \exists k \in \mathbb{N}^* : p^k \in J\}.$$

This is a 1:1 correspondence between algebaic varieties and radical ideals (根的理想) $\sqrt{J} \triangleleft R_n := F[x_1, \ldots, x_n]$.

Proof. First note that if $p^k(V) = 0$, then also $p(V) = 0 \in F$, because F is a field. Therefore $\sqrt{J} \subset I(V(J))$.

Next we need to check that $\sqrt{J}$ is indeed an ideal. Remember the homework where you showed that $\sqrt{(0)} \triangleleft R'$ is an ideal. If we define $\pi \colon R_n \twoheadrightarrow R' := R_n/J$ we see that $\pi(J) = (0)$. Then $\pi^{-1}\left(\sqrt{(0)}\right) = \sqrt{J}$ is the radical of J and thus an ideal.

The second part requires Hilbert's Nullstellensatz (希耳伯特零点定理). □

The radical ideals are also called *semiprime (半素的)*, because they are intersections of prime ideals.

We can also characterize irreducible varieties via their coordinate rings as follows:

Corollary 3.12.17. *An algebraic variety $V \subset \bar{F}^n$ is irreducible iff its coordinate ring $R(V)$ is a domain.*

Proof. Note that the vanishing ideal $I := I(V)$ is a radical ideal and therefore has no nilpotent elements. Now $I = I_1 \cap I_2$ with two strictly greater $I \subsetneq I_k$, iff $V = V(I_1) \cup V(I_2)$. But a radical ideal that is irreducible is prime. This completes the proof. □

Remark 3.12.18. The idea of algebraization of geometry is to assign rings to the spaces (赋环空间) under consideration and to discover the intrinsic geometric properties as invariants of the corresponding ring.

3.12.1 Algebraic dimension theory

The goal of this subsection is to show that the dimension of an algebraic variety can be reconstructed from its coordinate ring.

Example 3.12.19. 0. Consider the trivial ideal $(0) \triangleleft F[x_1, \ldots, x_n]$. Its corresponding variety is $V(0) = \bar{F}^n$ which is (an affine space) of dimension n. Correspondingly $F[x_1, \ldots, x_n]/(0) = F[x_1, \ldots, x_n]$ is a polynomial ring in n variables.

1. Conversely the other trivial ideal $I := F[x_1, \ldots, x_n]$ has an empty zero-locus $V(I) = \emptyset$ of dimension $-\infty$. Which corresponds to the trivial coordinate-"ring" $I/I = 0$.

2. The same is also true for other hyperplanes.

The full notion is called *Krull*[28]?? *dimension (克鲁尔维数)* of ideals.

Definition 3.12.20. *We say that the irreducible algebraic variety $V \subset \bar{F}^n$ has dimension d if the longest strictly decreasing sequence of varieties is $V = V_d \supsetneq V_{d-1} \supsetneq \cdots \supsetneq V_0 \supsetneq \emptyset$.*

The correspondence is now: $\dim V = n - \operatorname{hgt} I(V)$ where $\operatorname{hgt} I$ is the height or Krull dimension of a prime ideal in $F[x_1, \ldots, x_n]$.

3.12.2 Regular maps (常规映射)

Remember the coordinate ring $R(V) := F[x_1, \ldots, x_n]/I(V)$.

Proposition 3.12.21. *Given a finitely generated commutative associative unital algebra R over a field F with trivial nil-radical $\sqrt{(0)}$, then R is the coordinate ring of an algebraic variety over F.*

Proof. The universality property of $F[x_1, \ldots, x_n]$ implies that every commutative associative unital F-algebra R with n generators has a surjective homomorphism $\phi\colon F[x_1, \ldots, x_n] \twoheadrightarrow R$. But then $R \cong F[x_1, \ldots, x_n]/J$ where $J := \ker \phi \lhd F[x_1, \ldots, x_n]$. Since $\sqrt{(0)} = (0)$, we see that J is a radical ideal. But now $J = I(V(J))$ and thus $R \cong F[x_1, \ldots, x_n]/J = R(V(J))$. This completes the proof. □

Note that moreover $R(V)$ has the obvious *maximal ideals (极大理想)* $\mathfrak{m}_{\mathbf{x}} := \{f \in R(V) : f(\mathbf{x}) = 0\}$ for any $\mathbf{x} \in V$. (Maximality, because $R(V) \twoheadrightarrow F \cong R(V)/\mathfrak{m}_{\mathbf{x}} : f \mapsto f(\mathbf{x})$.)

Proposition 3.12.22. *Given an algebraic variety $V \subset \bar{F}^n$, then the maximal ideals $\mathfrak{m} \lhd R(V)$ correspond 1:1 to the points of V.*

This also follows from Hilbert's Nullstellensatz.
Note that this also implies:

Proposition 3.12.23. *Given a ring homomorphism $\phi\colon R(B) \to R(A)$, then there is a polynomial mapping $f\colon A \to B$ such that $\phi = f^*$.*

Proof. Note that ϕ induces a map of maximal ideals $f\colon \mathfrak{M}(A) \to \mathfrak{M}(B) : \mathfrak{m} \mapsto \phi^{-1}\mathfrak{m}$. Since the maximal ideals correspond to points in A and B, this means that there is a map $f\colon A \to B$. Given the coordinate functions $y_1, \ldots, y_n$ of $B \subset \bar{F}^n$, then these map to polynomials $f_k := \phi(y_k) \in R(A)$. Now for every $p \in R(B)$, the function $p \circ f\colon A \to \bar{F}$ is just $p \circ (f_1, \ldots, f_n)|A$, i.e. the map $f = (f_1, \ldots, f_n)$ a polynomial map. □

We therefore define:

Definition 3.12.24. *The* category (范畴) *of algebaic varieties over a field F is the family of algebraic varieties V over F together with the polynomial maps. An embedding of algebraic varieties is an injective polynomial map, an isomorphism of algebraic varieties is a pair of mutually inverse polynomial maps.*

The polynomial maps are sometimes also called *regular maps (常规映射)*. The last proposition implies now.

Corollary 3.12.25. *Given two algebraic varieties A, B over a field F, then they are isomorphic iff their corrdinate rings are isomorphic.*

Example 3.12.26. Consider the two algebraic varieties $A := V(y - x^2) \subset \bar{F}^2$ (the parabola) and $B := V(y) \subset \bar{F}^2$ (the x-axis). They are isomorphic via the maps $f(x, y) = (x, y - x^2)$ and $g(x, y) = (x, y + x^2)$ which map $f(A) = B$ and $g(B) = A$. This corresponds to the observation that their coordinate rings $R(A) = F[x, y]/(y - x^2) \cong F[x] = F[x, y]/(y) = R(B)$.

In this sense the study of algebraic varieties is equivalent to the study of their coordinate rings. This motivates the further study of ring theory.

3.12.99 Exercises

Exercise* 3.12.1. Let F be a field. Show that every proper ideal $I \lhd F[x_1, \ldots, x_n]$ (i.e. $I \neq F[x_1, \ldots, x_n]$) has a zero in $\bar{F}^n$.

Exercise 3.12.2. Show that the space $\bar{F}^n$ is compact (though not Hausdorff) in Zariski topology.

Exercise 3.12.3. Define the *spectrum* of a commutative ring R as $\operatorname{Spec} R := \{\mathfrak{p} \lhd R : \text{prime ideal}\}$. Show that for every ideal $I \lhd R$ of a coordinate ring of an algebraic variety, the set $V(I) := \{\mathfrak{p} \in \operatorname{Spec} R : I \subset \mathfrak{p}\}$ is closed in Zariski topology. *Hint:* You should first check that the Zariski topology on $\bar{F}^n$ induces a topology on $\operatorname{Spec} R$ which by abuse of notation is also called Zariski topology.

Exercise 3.12.4. Show that every regular mapping is continuous in Zariski topology.

Exercise 3.12.5. Let $f : B \to C$ and $g\colon A \to B$ be regular mappings of algebraic varieties over the same field F. Show that also $f \circ g$ is a regular mapping.

Remember the definition of localization $R_{\mathfrak{p}}$ of a ring by a prime ideal $\mathfrak{p} \lhd R$, i.e. the field of fractions $S^{-1}R$ for the multiplicative set $S := R \setminus \mathfrak{p}$.

Exercise 3.12.6. Let R be a domain and $F := K[R]$ its field of fractions. Show that for every $x \in F$, if $x \in R_{\mathfrak{m}} \subset F$ for every maximal ideal $\mathfrak{m} \lhd R$, then $x \in R$.

Exercise 3.12.7. Let $A \subset F^n$ be an irreducible algebraic variety over an algebraically closed field F. Define the rational functions on A as the set of functions $f\colon A \to F$ such that for every open neighborhood $U \subset A$, there is a rational function $p/q \in F(x_1, \ldots, x_n)$ with $f(\mathbf{x}) = p(\mathbf{x})/q(\mathbf{x})$ for all $\mathbf{x} \in U$. Show that the rational functions form a domain isomorphic to $R(A)$.

Chapter 4

Outlook: Category theory (范畴论, 2 weeks)

4.1 Categories and additive categories

4.1.1 Definition

Example 4.1.1. Remember vector spaces together with linear maps. The analogue for groups is (finite) groups together with group homomorphisms. The abstraction of the common properties of these two examples is a category.

Definition 4.1.2. *A* Category (范畴) $\mathcal{C}$ *is a family of* objects $\mathfrak{Obj}$ *together with a family of* morphisms $\mathfrak{Mor} = \bigcup_{A,B\in\mathfrak{Obj}} \mathfrak{Mor}(A,B)$ *and a composition* $\circ\colon \mathfrak{Mor}(B,C)\times \mathfrak{Mor}(A,B) \to \mathfrak{Mor}(A,C)$ *for every triple* $A,B,C \in \mathfrak{Obj}$ *subject to the following axioms*

1. *composition of morphisms is associative, i.e. for all* $f \in \mathfrak{Mor}(C,D)$, $g \in \mathfrak{Mor}(B,C)$, *and* $h \in \mathfrak{Mor}(A,B)$, *then* $f\circ(g\circ h) = (f\circ g)\circ h$.
2. *For every object* $A \in \mathfrak{Obj}$ *there is a morphism* $\mathrm{id}_A \in \mathfrak{Mor}(A,A)$ *such that for all* $f \in \mathfrak{Mor}(A,B)$, $f\circ \mathrm{id}_A = f$ *and for all* $g \in \mathfrak{Mor}(B,A)$, $\mathrm{id}_A \circ g = g$.

Given a morphism $f \in \mathfrak{Mor}(A,B)$ *together with a morphism* $g \in \mathfrak{Mor}(B,A)$ *such that* $f\circ g = \mathrm{id}_B$ *and* $g\circ f = \mathrm{id}_A$ *then we call* f *and* g equivalences *and* A *and* B equivalent.

Example 4.1.3.

1. The category $\mathfrak{Set}$ consists of sets as objects together with maps (i.e. functions) as morphisms.
2. The category $\mathfrak{Gp}$ consists of groups as objects and group homomorphisms as morphisms. Remember that composition of homomorphisms is associative and the identity is $\mathrm{Id}_G\colon G \to G : g \mapsto g$.

category	groups $\mathfrak{Gp}$	$\supset$	abelian groups $\mathfrak{Ab}$	rings $\mathfrak{Ri}$	$\supset$	fields $\mathfrak{Fi}$
objects	group $(G, \cdot)$		ab. group $(A, +)$	ring $(R, +, \cdot)$		field $(F, +, \cdot)$
morphism	group homomorphism		ab gr. hom	ring homomorphism		field homomorphism
monomorph.	embedding		embedding	embedding		all
epimorph.	projection		projection	projection		iso
isomorph.	group iso.		ab. gr.iso	ring iso		field iso
kernels	normal subgroup		ab. subgroup	ideal		trivial
1st iso thm/ hom-thm			$\operatorname{im}\phi \cong D(\phi)/\ker\phi$			trivial
2nd iso thm	$(G/K)/(N/K) \cong G/N$ for $K, N \lhd G$, $K \subset N$			$(R/I)/(J/I) \cong R/J$ for $I, J \lhd R$, $I \subset J$		trivial
3rd iso thm	$(SN)/N \cong S/(S \cap N)$ for $S \subset G$, $N \lhd G$			$(S + I)/I \cong S/(S \cap I)$ $S \subset R$, $I \lhd R$		trivial

$\mathfrak{Ab} \subset \mathfrak{Gp}$ is a *full subcategory*, i.e. abelian groups are special groups (a sub-"set") and a group hom between abelian groups is automatically an abelian group hom.

$\mathfrak{Fi} \subset \mathfrak{Ri}$ is also a full subcategory. But $\mathfrak{Ri}$ is *not* a subcategory of $\mathfrak{Ab}$, because rings contain additional data. (E.g. $C_2 \times C_2$ can be made into a ring in two ways: $(\mathbb{Z}/(2)) \times (\mathbb{Z}/(2))$ or $\mathbb{F}_4$.)

A *functor* is a map between two categories, e.g. Galois correspondence Gal: $\mathfrak{Gal} \to \mathfrak{Gp}$ where $\mathfrak{Gal}$ are Galois extensions and $\mathrm{Gal}(E : F)$ is the Galois group of the extension. (Note that this functor is *contra-variant*, i.e. $\mathrm{Gal}(E : F \subset E : K) = \mathrm{Gal}(E : F) \supset \mathrm{Gal}(E : K)$.)

3. The *subcategory (范畴子)* $\mathfrak{Ab}$ consists of abelian groups and their group homomorphisms. Since for every pair of abelian groups it contains all their homomorphisms, it is called a *full* subcategory.

4. The category $\mathfrak{Var}_F$ consists of algebraic varieties over the field F and regular maps between them.

5. The category $\mathfrak{Vect}_F$ consists of (finite dimensional) vector spaces over the field F. It splits into the subcategories $\mathfrak{Vect}_n$ of n-dimensional vector spaces over F.

6. Given a category $\mathcal{C}$, we denote by $\mathcal{C}^{op}$ the category with $\mathfrak{Obj}_{\mathcal{C}^{op}} := \mathfrak{Obj}_{\mathcal{C}}$, $\mathfrak{Mor}_{\mathcal{C}^{op}}(A,B) := \mathfrak{Mor}_{\mathcal{C}}(B,A)$ for all $A, B \in \mathfrak{Obj}_{\mathcal{C}^{op}} = \mathfrak{Obj}_{\mathcal{C}}$ and $f \circ_{\mathcal{C}^{op}} g := g \circ f$ for all $f \in \mathfrak{Mor}_{\mathcal{C}^{op}}(B,C) = \mathfrak{Mor}_{\mathcal{C}}(C,B)$, $g \in \mathfrak{Mor}_{\mathcal{C}^{op}}(A,B) = \mathfrak{Mor}_{\mathcal{C}}(B,A)$. It is easy to see that $\mathcal{C}^{op}$ also fulfills the axioms of a category. The difference to $\mathcal{C}$ is that morphisms compose in the opposite order.

4.1.2 Functor

Definition 4.1.4. *Given two categories $\mathcal{C}$ and $\mathcal{D}$. A* (covariant) functor (函子) *$F\colon \mathcal{C} \to \mathcal{D}$ is a map $F\colon \mathfrak{Obj}_{\mathcal{C}} \to \mathfrak{Obj}_{\mathcal{D}}$ together with maps $F_{A,B}\colon \mathfrak{Mor}_{\mathcal{C}}(A,B) \to \mathfrak{Mor}_{\mathcal{D}}(F(A),F(B))$ for all $A, B \in \mathfrak{Obj}_{\mathcal{C}}$, subject to the rules $\forall A, B, C \in \mathfrak{Obj}_{\mathcal{C}}$ and $f \in \mathfrak{Mor}_{\mathcal{C}}(B,C), g \in \mathfrak{Mor}_{\mathcal{C}}(A,B)$*

$$F_{A,A}(\mathrm{id}_A) = \mathrm{id}_{F(A)}, \tag{4.1}$$
$$F_{A,C}(f \circ g) = F_{B,C}(f) \circ F_{A,B}(g). \tag{4.2}$$

A contravariant functor (反变函子) *$F\colon \mathcal{C} \to \mathcal{D}$ is a (covariant) functor $F\colon \mathcal{C}^{op} \to \mathcal{D}$.*

Example 4.1.5. 1. Consider the map $\mathrm{ab}\colon \mathfrak{Gp} \to \mathfrak{Ab}$ that sends a group G to its abelization $\mathrm{ab}(G) := G/[G,G]$ where $[G,G] := \langle [g,h] := ghg^{-1}h^{-1} : g,h \in G\rangle$ is the commutator subgroup, i.e. the smallest normal subgroup that contains all commutators in G. For a group homomorphism $f\colon G \to H$ we assign the map $\mathrm{ab}_{G,H}(f) := \bar{f}\colon G/[G,G] \to H/[H,H] : g[G,G] \mapsto f(g)[H,H]$. Note that the homomorphism property implies $f[g_1,g_2] = [f(g_1),f(g_2)]$, i.e. $f[G,G] \subset [H,H]$ and thus $\bar{f}$ is well-defined. Finally $\mathrm{ab}(f\circ f')(g[G,G]) = (f\circ f')(g)[K,K] = f(f'(g))[K,K] = \mathrm{ab}(f)(f'(g)[H,H]) = \mathrm{ab}(f)(\mathrm{ab}(f')(g[G,G]) = (\mathrm{ab}(f)\circ\mathrm{ab}(f'))(g[G,G])$ for $f \in \mathrm{Hom}(H,K)$, $f' \in \mathrm{Hom}(G,H)$, and $g \in G$. Therefore $\mathrm{ab}(f\circ f') = \mathrm{ab}(f)\circ\mathrm{ab}(f')$. By analog computations also $\mathrm{ab}(\mathrm{id}_G) = \mathrm{id}_{\mathrm{ab}(G)}$, i.e. identities are mapped to identities.

2. Given a subcategory $\mathcal{D} \subset \mathcal{C}$, then the embedding $e\colon \mathcal{D} \to \mathcal{C}$ is a functor. In particular the identity of any category $\mathcal{C}$, $\mathrm{Id}_{\mathcal{C}}\colon \mathcal{C} \to \mathcal{C}$ is a functor.

3. If we consider the category $\mathbb{Gal}_p$ of (finite) Galois extensions of fields with characteristic $p \in \mathbb{P}$ (or 0), then $\mathrm{Gal}\colon \mathbb{Gal}_p \to \mathfrak{Gp}$ is a contravariant functor, i.e. to every Galois extension it assigns its Galois group, to a morphism of Galois extensions, i.e. $f\colon (F \subset E) \to (K \subset E)$ it assigns a morphism of groups $\mathrm{Gal}(f)\colon \mathrm{Gal}(E : K) \to \mathrm{Gal}(E : F)$ (in the opposite direction), and then the composition is contravariant, i.e. $\mathrm{Gal}(f \circ g) = \mathrm{Gal}(g) \circ \mathrm{Gal}(f)$ for $g\colon (F_0 \subset E) \to (F \subset E)$ with $F_0 \subset F$.

4.1.3 Mono-, Epi- and Isomorphism

What about the notion of injective and surjective?

Lemma 4.1.6. *Given a* concrete category*, i.e. all $A \in \mathfrak{Obj}$ are sets (with additional structure) and all $f \in \mathfrak{Mor}(A, B)$ are maps of sets (for all $A, B \in \mathfrak{Obj}$). Then $f \circ g = \mathrm{Id}_A$ for $f \in \mathfrak{Mor}(B, A)$ and $g \in \mathfrak{Mor}(A, B)$ implies that f is surjective and g is injective.*

This follows easily from elementary set theory.

Unfortunately the opposite is in general wrong, i.e. $f \in \mathfrak{Mor}(A, B)$ with $f\colon A \to B$ injective does not imply that there is any morphism $g \in \mathfrak{Mor}(B, A)$.

If there are $f \in \mathfrak{Mor}(A, B)$, $g \in \mathfrak{Mor}(B, A)$ with $f \circ g = \mathrm{id}_A$ and $h_i \in \mathfrak{Mor}(C, B)$, then $g \circ h_1 = g \circ h_2$ implies $h_1 = h_2$ by composing with f from the left. Conversely for $h_i \in \mathfrak{Mor}(B, C)$, $h_1 \circ f = h_1 \circ f$ implies $h_1 = h_2$ by composing with g from the right.

Therefore we define:

Definition 4.1.7. *Given a category $\mathcal{C}$. We say that $f \in \mathfrak{Mor}(A, B)$ is a* monomorphism *if for every $C \in \mathfrak{Obj}$ and every pair $g_i \in \mathfrak{Mor}(C, A)$, $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$.*

We say that f is an epimorphism *if for every pair $h_i \in \mathfrak{Mor}(B, C)$, $h_1 \circ f = h_2 \circ f$ implies $h_1 = h_2$.*

Therefore isomorphisms are monomorphisms and epimorphisms. Unfortunately a morphism $f \in \mathfrak{Mor}(A, B)$ being monomorphism and epimorphism does not imply that there is any morphism in $\mathfrak{Mor}(B, A)$, so we cannot conclude that it is an isomorphism. We could call f in this situation *quasi-isomorphism*.

Proposition 4.1.8. *Given a group-homomorphism, then it is a monomorphism iff it is injective, i.e.* $\ker f = 1$ *for the homomorphism f.*

Proof. The back-direction is obvious. Given thus a monomorphism $f\colon G \to H$. Assume $f(x) = f(y)$ for some $x, y \in G$. Since $(\mathbb{Z}, +)$ is generated by 1, we can define homomorphisms by letting $g_1\colon 1 \mapsto x$ and $g_2\colon 1 \mapsto y$ and extending as group homomorphisms $\mathbb{Z} \to G$. Because f is a monomorphism and $f \circ g_1 = f \circ g_2$ we know $g_1 = g_2$ and thus $x = g_1(1) = g_2(1) = y$. □

The same property carries over to rings.

Proposition 4.1.9. *Given a group homomorphism. It is an epimorphism iff it is surjective.*

Proof. Again the back direction is obvious. Conversely assume that $f\colon G \to H$ is a group homomorphism that is not surjective. Therefore we can construct isomorphic copies $H_k \cong H$ that are coinciding as sets in $\operatorname{im} f$ which is a subgroup, i.e. $H_1 \cap H_2 = \operatorname{im} f$ (e.g. $\operatorname{im} f$ and H). Consider the free product with amalgamation $P := H_1 * H_2 / \operatorname{im} f$, i.e. in the free product $F := H_1 * H_2$ (considered as disjoint groups) we divide out the normal subgroup generated by $\langle i_1(x) i_2(x)^{-1} : x \in \operatorname{im} f \rangle$ for $i_k\colon H_k \to F$. Now the embeddings $h_k\colon H \cong H_k \to P$ are different, because for the elements $H_k \setminus \operatorname{im} f$ they differ, but $h_1 \circ f = h_2 \circ f$. Therefore f cannot be an epimorphism. □

A similar result holds for vector spaces, but not for rings or R-algebras (see Exercise **??**).

4.1.4 Initial and Terminal Objects

Definition 4.1.10. *Given a category* $\mathcal{C}$*, then an* initial object $\emptyset$ *is an object* $\emptyset \in \mathfrak{Obj}$ *such that for every object* $A \in \mathfrak{Obj}$ *there is exactly one morphism* $\{\emptyset\} = \mathfrak{Mor}(\emptyset, A)$.

An object pt *is called* terminal object *if for every object* $A \in \mathfrak{Obj}$ *there is a unique morphism* $\{\mathrm{pt}\} = \mathfrak{Mor}(A, \mathrm{pt})$.

If an object $0 \in \mathfrak{Obj}$ *is initial and terminal object, then it is called a* null object.

Example 4.1.11.

1. In the category $\mathfrak{Set}$, $\emptyset$ is an initial object and $\mathrm{pt} = \{*\}$ a set consisting of only one element is a terminal object. It is easy to show that initial (terminal) object are unique (up to isomorphism if they exist). But $\emptyset \neq \{*\}$. Therefore this category has no null object.

2. In the category of groups however the initial object is $1 := \{\mathrm{id}\}$ the trivial group. Which is at the same time also the terminal object and therefore also the null object. The same object also works for abelian groups where it is denoted 0, as well as for vector spaces over a fixed field F.

4.1.99 Exercises

Exercise 4.1.1. A groupoid is a small category in which every morphism is an isomorphism. Let M denote the objects of a groupoid and $G(a,b)$ the morphisms for $a,b \in M$. Let further $G_a := G(a,a)$.

a. Show that G_a is a group;

b. Assume that there is a morphism $\gamma \in G(a,b)$. Show that G_a and G_b are conjugate.

Note the similarity to a group action.

Exercise 4.1.2. Given a category $\mathcal{C}$, denote $\mathcal{C}^{op}$ the category with the same objects and morphisms, but $\mathfrak{Mor}_{op}(A,B) := \mathfrak{Mor}(B,A)$ and composition correspondingly, i.e. $\circ_{op} \colon \mathfrak{Mor}_{op}(B,C) \times \mathfrak{Mor}_{op}(A,B) \to \mathfrak{Mor}_{op}(A,C) : (f,g) \mapsto g \circ f$.

a. Show that $\mathcal{C}^{op}$ is a category iff $\mathcal{C}$ is a category.

b. What happens to initial/ terminal objects?

c. What happens to epi- / monomorphisms?

Exercise 4.1.3. Remember that a partially ordered set $(S, \leq)$ is a set together with a reflexive, anti-symmetric (i.e. $x \leq y$ and $y \leq x$ then $x = y$), transitive relation.

a. Show that we obtain a small category by setting $\mathfrak{Obj} = S$ and $\mathfrak{Mor} = |\leq|$, i.e. there is exactly one morphism $f \colon x \to y$ iff $x \leq y$.

b. What is the condition for a small category to be generated by a partially ordered set?

c. What are its mono-, epi- and isomorphisms?

Exercise 4.1.4 (Free Category)**.** Remember that an oriented multi graph G is a set V (the vertices), a set E (the edges) together with two maps $b \colon E \to V$ and $\mathrm{end} \colon E \to V$ (called beginning and end of edges). A path of length $n \in \mathbb{N}$ is a sequence $v_0 \xrightarrow{e_1} v_1 \xrightarrow{e_2} v_2 \to \ldots v_{n-1} \xrightarrow{e_n} v_n$ such that $b(e_k) = v_{k-1}$ and $\mathrm{end}(e_k) = v_k$ for all k. Paths are composed by concatenation.

Show that the vertices together with the paths of a graph under composition give a small category.

Exercise 4.1.5 (Product of Categories)**.** Given two categories $\mathcal{C}$ and $\mathcal{D}$, then their product $\mathcal{C} \times \mathcal{D}$ is the category whose objects are $\mathfrak{Obj} := \mathfrak{Obj}_{\mathcal{C}} \times \mathfrak{Obj}_{\mathcal{D}}$ pairs of an object from $\mathcal{C}$ and an object from $\mathcal{D}$. The Morphisms are $\mathfrak{Mor} := \mathfrak{Mor}_{\mathcal{C}} \times \mathfrak{Mor}_{\mathcal{D}}$ pairs of morphisms from $\mathcal{C}$ and $\mathcal{D}$.

a. Show that this is indeed a category. What is the composition law?

b. Define the notion of bifunctor, e.g. $\otimes$ on $\mathfrak{Vect}$.

c. Show that $\mathrm{Hom}_{\mathcal{C}}(.,.)\colon \mathcal{C}^{op} \times \mathcal{C} \to \mathfrak{Set}$ for a locally small category $\mathcal{C}$ (i.e. the morphisms $\mathfrak{Mor}(.,.)$ form a set) with $\mathrm{Hom}_{\mathcal{C}}(A,B) = \mathfrak{Mor}(A,B)$ for $A, B, C, D \in \mathfrak{Obj}$ and $\alpha \in \mathfrak{Mor}(A,B)$, $\gamma \in \mathfrak{Mor}(C,D)$, then

$$\mathrm{Hom}_{\mathcal{C}}(\alpha,\gamma)\colon \mathrm{Hom}_{\mathcal{C}}(B,C) \to \mathrm{Hom}_{\mathcal{C}}(A,D) : \beta \mapsto \alpha \circ \beta \circ \gamma$$

is a bifunctor.

Exercise 4.1.6 (Yoneda's[1] Lemma)**.** Given a locally small category $\mathcal{C}$ together with a functor $F\colon \mathcal{C} \to \mathfrak{Set}$ and using the definition of $\mathrm{Hom}_{\mathcal{C}}(.,.)$ from the previous exercise, show that for each object $A \in \mathcal{C}$ there is a 1:1-correspondence between elements of $F(A)$ and natural transformations from $\mathrm{Hom}_{\mathcal{C}}(A,\bullet)$ to $F(\bullet)$.
Hint: Does it help to see for a natural transformation $\tau_\bullet\colon \mathrm{Hom}_{\mathcal{C}}(A,\bullet) \to F(\bullet)$ there is the element $\tau_A(\mathrm{id}_A) \in F(A)$.

4.2 Limits and Colimits

4.2.1 Products and Coproducts

Remember again the category of abelian groups. Beside groups, homomorphisms, and composition of homomorphisms, we also have another interesting operation: Direct product. Remember (see e.g. Exercise 1.5.1-a) that the direct product of two groups G_i induces the maps $\pi_i\colon G_1 \times G_2 \to G_i$ such that for every pair of maps $\phi_i\colon H \to G_i$ there is a unique map $(\tilde{\phi}\colon H \to G_1 \times G_2) = (\phi_1, \phi_2)$. This can be generalized to the following definition:

Definition 4.2.1. *Given a category $\mathcal{C}$ and two objects $G_i \in \mathfrak{Obj}$. A product is an object P together with two epimorphisms $p_i \in \mathfrak{Mor}(P, G_i)$ such that for every pair $\phi_i \in \mathfrak{Mor}(H, G_i)$ there is a unique map $\tilde{\phi} \in \mathfrak{Mor}(H, P)$ with $\phi_i = p_i \circ \tilde{\phi}$, $i = 1, 2$.*

Due to its characteristics the product (if it exists) is unique up to isomorphism. We denote it as $G_1 \times G_2$, or $\prod_{i \in I} A_i$ for more factors.

For every construction in categories there is the dual construction (which you can think of as the same construction in the dual category, see Exercise 4.1.2) where the arrows are reverted. The dual to a product is a coproduct as follows:

[1] Nobuo Yoneda *1930/3 in Japan, †1996/4

Definition 4.2.2. *Given a category $\mathcal{C}$ and two objects $G_i \in \mathfrak{Obj}$. A coproduct is an Object $C \in \mathfrak{Obj}$ together with two monomorphisms $e_i \in \mathfrak{Mor}(G_i, C)$ such that for every pair $\epsilon_i \in \mathfrak{Mor}(G_i, H)$ there is a unique map $\tilde{\epsilon} \in \mathfrak{Mor}(C, H)$ with $\epsilon_i = \tilde{\epsilon} \circ e_i$, $i = 1, 2$.*

Again, due to its characteristics the coproduct (if it exists) is unique up to isomorphism. In the category of sets it is the disjoint union and is thus denoted as $G_1 \sqcup G_2$ or for more objects $\coprod_{i \in I} A_i$.

4.2.2 Equalizer and Coequalizer

Definition 4.2.3. *Given a category $\mathcal{C}$ and two morphisms $f, g \colon A \rightrightarrows B$ for objects $A, B \in \mathfrak{Obj}$, an* equalizer *(if it exists) is a morphism $e \colon E \to A$ for some object $E \in \mathfrak{Obj}$ such that $f \circ e = g \circ e$ and for every morphism $\phi \colon C \to A$ with $f \circ \phi = g \circ \phi$ there is a unique $\bar{\phi} \colon C \to E$ with $\phi = e \circ \bar{\phi}$.*

In the category of groups, the equalizer between $f \colon G \to H$ and $0 \colon G \to H : g \mapsto \mathrm{id}_H$ is the inclusion $e \colon \ker f \hookrightarrow G$.

It is easy to see that an equalizer is a monomorphism (because $\bar{\phi}$ is unique).

The dual construction is

Definition 4.2.4. *Given a category $\mathcal{C}$ and two morphisms $f, g \colon A \rightrightarrows B$ for objects $A, B \in \mathfrak{Obj}$, a* coequalizer *(if it exists) is a morphism $p \colon B \to C$ such that $p \circ f = p \circ g$ and for every morphism $\phi \colon B \to D$ with $\phi \circ f = \phi \circ g$ there is a unique $\bar{\phi} \colon C \to D$ with $\phi = \bar{\phi} \circ p$.*

In the category of groups, the coequalizer between $f \colon G \to H$ for $\operatorname{im} f \lhd H$ and $0 \colon G \to H : g \mapsto \mathrm{id}_H$ is the projection $p \colon H \to H / \operatorname{im} f$. For arbitrary homomorphism f, we call $\operatorname{coker} f := H / \operatorname{im} f$ the *cokernel* of f.

4.2.3 Limts

Definition 4.2.5. *A* diagram *in a category $\mathcal{C}$ over a (small) directed graph G is a functor from the free category $\hat{G}$ to $\mathcal{C}$.*

A diagram $\mathcal{D}$ in $\mathcal{C}$ is said to be commuting *if for every pair of paths $p, q \colon v_0 \to v_n$ the morphisms $\mathcal{D}(p) = \mathcal{D}(q)$ are equal.*

Due to the freeness of $\hat{G}$, the functor is encoded by two compatible maps $\mathcal{D}_\bullet \colon V \to \mathfrak{Obj}$ and $\mathcal{D}(\bullet) \colon E \to \mathfrak{Mor}$.

Definition 4.2.6. *Let $A \in \mathfrak{Obj}$ be an object in a category $\mathcal{C}$ and let $\mathcal{D}$ be a diagram in $\mathcal{C}$ over a graph G. A cone from A to $\mathcal{D}$ assigns to each vertex $v \in V(G)$ a morphism $\phi_v \colon A \to \mathcal{D}_v$ such that $\mathcal{D}(e) \circ \phi_v = \phi_w$ for every edge $e \colon v \to w$.*

A cone from A to $\mathcal{D}$ is equivalent to a morphism from the constant diagram $C(A)$ over G to the diagram $\mathcal{D}$. Because morphisms of diagrams over the same graph are composable, we have the following two properties:

1. If $\phi\colon A \to \mathcal{D}$ is a cone and $\psi\colon \mathcal{D} \to \mathcal{E}$ a morphism of diagrams, then $\psi \circ \phi\colon A \to \mathcal{E}$ is a cone.

2. If $\phi\colon B \to \mathcal{D}$ is a cone and $\psi\colon A \to B$ is a morphism, then $\phi \circ \psi\colon A \to \mathcal{D}$ is a cone.

We are now ready to define limits:

Definition 4.2.7. *Let $\mathcal{D}$ be a diagram in a category $\mathcal{C}$ over a graph G. A* limit cone *of D is a cone $\lambda\colon L \to \mathcal{D}$ such that, for every cone $\phi\colon A \to \mathcal{D}$ there is a unique morphism $\bar{\phi}\colon A \to L$ with $\phi = \lambda \circ \bar{\phi}$.*

Limits are unique in the following sense

Proposition 4.2.8. *Given two limit cones $\lambda^{(i)}\colon L_i \to \mathcal{D}$, then there is an isomorphism $\tau\colon L_1 \to L_2$ such that $\lambda^{(1)} = \lambda^{(2)} \circ \tau$. Conversely, given a limit cone $\lambda\colon L \to \mathcal{D}$ and an isomorphism $\tau\colon L' \to L$, then $(\lambda'\colon L' \to \mathcal{D}) := \lambda \circ \tau$ is a limit cone.*

Example 4.2.9. 0. A direct product is the limit cone of a diagram over a discrete graph, i.e. one without edges.

1. The equalizer is the limit cone of the diagram with two parallel morphisms.

2. A *pullback* of a morphism $f\colon A \to C$ along a morphism $g\colon B \to C$ is a limit cone to the following diagram

$$\begin{array}{ccc} & & A \\ & & \downarrow{\scriptstyle f} \\ B & \xrightarrow{g} & C \end{array}$$

3. The *projective limit* is the limit of a diagram whose graph is a ray coming from infinity:

$$\cdots \to D_2 \to D_1 \to D_0$$

4.2.4 Colimits

The dual notion is that of a colimit.

Definition 4.2.10. *A colimit cone of a diagram $\mathcal{D}$ over a graph G is a limit cone of the same diagram in the category $\mathcal{C}^{op}$ over the graph G^{op}.*

The "cocones" are the morphisms $\phi_\bullet \colon \mathcal{D} \to C(A)$ for some object $A \in \mathfrak{Obj}$ and $C(A)$ denoting the constant diagram over the same graph.

Example 4.2.11. 0. the coproduct is the colimit cone of a diagram over a singular graph (without edges).

1. The coequalizer is the colimit cone of the diagram with two parallel morphisms.

2. The *push out* of two morphisms $f \colon A \to B$ and $g \colon A \to C$ is the colimit cone of the diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle g}\downarrow & & \\ C & & \end{array}$$

3. The *inductive limit* is the colimit of a diagram whose graph is a ray going to infinity:

$$D_0 \to D_1 \to D_2 \to \dots$$

4.2.5 Construction of Limits and Colimits

Given an arbitrary diagram, you may wonder whether its limit/ colimit exists. This is captured by the following property.

Definition 4.2.12. *A category $\mathcal{C}$ is (co)complete if every diagram has a (co)limit.*

As long as we talk about concrete categories (i.e. the objects are sets with additional structure), there is a chance that the category is complete, because:

Proposition 4.2.13. *The category $\mathfrak{Set}$ is (co)complete.*

Proof. Let $\mathcal{D}$ be a diagram in $\mathfrak{Set}$ over a graph G. Let $P := \prod_{v \in V(G)} \mathcal{D}_v$ be the carthesian product of all sets in $\mathcal{D}$ with projections $\pi_v \colon P \to \mathcal{D}_v$. We show that

$$L := \{(x_v : v \in V(G)) : \forall (e : v \to w) \in E(G) : \mathcal{D}(e)x_v = x_w\}$$

is a limit of $\mathcal{D}$ with limit cone $\lambda_v = \pi_v | L$. By inspection λ is a cone. If $\phi \colon S \to \mathcal{D}$ is a cone, then $\mathcal{D}(e)(\phi_v(x)) = \phi_w(x)$ for all $e \colon v \to w$ and $x \in A$. Hence $\bar{\phi} \colon A \to L$:

$x \mapsto (\phi_v(x))_{v \in V(G)}$ defines a morphism with the property $\phi = \lambda \circ \bar{\phi}$. By inspection this is the only such morphism. Therefore λ is a limiting cone.

The proof of cocompleteness is analogous and left as an exercise. □

If we inspect the construction, we see how to generalize this to arbitrary categories:

Proposition 4.2.14. *A category that has small products and equalizers is complete.*

Proof. Let $\mathcal{D}$ be a diagram in $\mathcal{C}$ over a graph G. Start with the direct product $P := \prod_{v \in V(G)} \mathcal{D}_v$ and for the end points $d(e\colon v \to w) = w$ define $Q := \prod_{e \in E(G)} \mathcal{D}_{d(e)}$ with projections $\pi_v\colon P \to \mathcal{D}_v$ and $\rho_e\colon Q \to \mathcal{D}_{d(e)}$. The universal property of Q yields unique morphisms $\alpha, \beta\colon P \rightrightarrows Q$ with $\rho_e \circ \alpha = \pi_{d(a)}$ and $\rho_a \circ \beta = \pi_{b(e)}$ where $b(e\colon v \to w) = v$ the beginning of the edge. We can now express the filter condition $\mathcal{D}(e)x_v = x_w$ for the set construction as $\alpha(x) = \beta(x)$. In terms of categories this is exactly the equalizer, i.e. let $\epsilon\colon E \to P$ be the equalizer of α and β in $\mathcal{C}$, and $\lambda_v := \pi_v \circ \epsilon$. We show that $\lambda = (\lambda_v)$ is a limit cone of $\mathcal{D}$. If $e\colon v \to w$ is an edge in G, then $\rho_e \circ \alpha = \pi_v$ and $\rho_e \circ \beta = \mathcal{D}(e) \circ \pi_w$, and so $\mathcal{D}(e) \circ \lambda_v = \mathcal{D}(e) \circ \pi_v \circ \epsilon = \rho_e \circ \beta \circ \epsilon = \rho_e \circ \alpha \circ \epsilon = \pi_w \circ \epsilon = \lambda_w$. Thus $\lambda\colon L \to \mathcal{D}$ is a cone. Now, let $\phi\colon A \to \mathcal{D}$ be any cone. Let $\tilde{\phi}\colon A \to P$ the unique morphism with $\pi_v \circ \tilde{\phi} = \phi_v$ for all $v \in V(G)$. Because ϕ is a cone, $\rho_e \circ \alpha \circ \tilde{\phi} = \pi_w \circ \tilde{\phi} = \phi_w = \mathcal{D}(e) \circ \phi_v = \mathcal{D}(e) \circ \phi_v = \mathcal{D}(e) \circ \pi_v \circ \tilde{\phi} = \rho_e \circ \beta \circ \tilde{\phi}$ for every edge $e\colon v \to w$ in G. Therefore $\alpha\tilde{\phi} = \beta\tilde{\phi}$ and there is a unique morphism $\bar{\phi}\colon A \to L$ such that $\tilde{\phi} = \epsilon \circ \bar{\phi}$. Thus $\bar{\phi}$ is unique with $\phi_v = \lambda_v \circ \bar{\phi}$ for all $v \in V(G)$. □

Corollary 4.2.15. *The categories* $\mathfrak{Gp}$, $\mathfrak{Vect}$ *and* $\mathfrak{Ri}$ *are complete.*

This follows from the construction of products and equalizers in these categories.

Remember that a category is cocomplete iff its opposite is complete. The dual of the above construction gives a cocompleteness condition as follows:

Proposition 4.2.16. *Given a category that has coproducts and coequalizers, then it is cocomplete.*

The details of the proof are left as an exercise.

The difficulty with cocompleteness is not as much the construction of coproducts, but the existence of coequalizers. Therefore $\mathfrak{Gp}$ is not cocomplete, but $\mathfrak{Vect}$ is.

Definition 4.2.17. *A graph is* finite *if it has finitely many vertices and finitely many edges. A finite limit is a limit of a diagram over a finite graph.*

Proposition 4.2.18. *A category that has equalizers and finite (co)products has finite (co)limits.*

This proof is by induction over the vertices and edges (in the construction of P and Q in the proof for completeness).

4.2.6 Functoriality of (Co)Limits

Let $\lambda\colon L \to \mathcal{D}$ and $\lambda'\colon L' \to \mathcal{D}'$ be limit cones of diagrams in the same category over the same graph G. If $\alpha\colon \mathcal{D} \to \mathcal{D}'$ is a morphism of diagrams, then $\alpha \circ \lambda$ is a cone to $\mathcal{D}'$ and there is a unique morphism $\lim \alpha\colon L \to L'$ such that $\alpha_v \circ \lambda_v = \lambda'_v \circ (\lim \alpha)$ for all $v \in V(G)$.

It is immediate that $\lim \mathrm{id}_{\mathcal{D}} = \mathrm{id}_{\lim \mathcal{D}}$ and that $\lim(\alpha \circ \beta) = (\lim \alpha) \circ (\lim \beta)$. Hence in a complete category, we have a functor $\lim\colon \mathfrak{Diag}(G, \mathcal{C}) \to \mathcal{C}$. The analogue is true for colimits in cocomplete categories.

Due to this functoriality (co)limits (and in particular (co)products) are compatible with the Hom-functor in the following sense:

Proposition 4.2.19. *Given a category with internal* Hom*-objects. Then the functor* $\mathrm{Hom}_{\mathcal{C}}(A, \bullet)$ *preserves existing limits.*

Dually $\mathrm{Hom}_{\mathcal{C}}(\bullet, B)$ *changes colimits into limits (if they exist).*

Example 4.2.20. For group-homomorphism this specializes to the known properties

$$\mathrm{Hom}_{\mathfrak{Gp}}(A, \prod_{i \in I} B_i) \cong \prod_{i \in I} \mathrm{Hom}(A, B_i)$$

and

$$\mathrm{Hom}_{\mathfrak{Gp}}(\coprod_{i \in I} A_i, B) \cong \prod_{i \in I} \mathrm{Hom}(A_i, B).$$

4.2.7 Additive and Abelian Categories

Definition 4.2.21.

4.2.99 Exercises

Exercise 4.2.1 (Cocompleteness of $\mathfrak{Set}$)**.** Show that the category $\mathfrak{Set}$ is cocomplete, i.e. every diagram in $\mathfrak{Set}$ has a colimit cone.
Hint: Use the construction dual to the one in the completeness proof of Proposition 4.2.13.

Exercise 4.2.2. Show Proposition 4.2.16, i.e. a category that has (small) coproducts and coequalizers is cocomplete.

Exercise 4.2.3. Show that the forgetful functor from $\mathfrak{Gp}$ to $\mathfrak{Set}$ preserves limits.

Exercise 4.2.4. Show that the forgetful functor F from $\mathfrak{Gp}$ to $\mathfrak{Set}$ creates limits, i.e. if $\mathcal{D}\colon G \to \mathfrak{Gp}$ is a diagram and $\mu\colon M \to F \circ \mathcal{D}$ is a limit cone in $\mathfrak{Set}$, then there is a unique cone $\lambda\colon L \to \mathcal{D}$ in $\mathfrak{Gp}$ with $F(\lambda) = \mu$ and it is a limit cone.

Exercise 4.2.5. Show that a category is complete iff it has products and pullbacks.

Exercise 4.2.6.

4.3 Tensor products: tensor algebra, symmetric algebra, exterior algebra

Definition 4.3.1. *Given an additive category $\mathcal{C}$, then a tensor product is a map $\otimes\colon \mathfrak{Obj}\times\mathfrak{Obj} \to \mathfrak{Obj}$ together with a map $\otimes\colon \mathfrak{Mor}(A,B)\times\mathfrak{Mor}(C,D) \to \mathfrak{Mor}(A\otimes C, B\otimes D)$ such that ....*

Example 4.3.2. 1. Given the category $\mathfrak{Vect}_F$ of (finite dimensional) vector spaces over a field F, then $V \otimes W$ is the vector space $\langle V \times W\rangle_F/\!\sim$ where we identify

$$\begin{aligned}\lambda(v,w) &\sim (\lambda v, w) \sim (v, \lambda w),\\ (v_1+v_2, w) &\sim (v_1,w)+(v_2,w),\\ (v, w_1+w_2) &\sim (v,w_1)+(v,w_2)\end{aligned}$$

for all $v_i \in V$, $w_i \in W$ and $\lambda \in F$. It can be shown that $V\otimes W$ is finite dimensional if V and W are, more precisely $\dim_F(V\otimes W) = (\dim_F V)(\dim_F W)$ and in particular $V\otimes 0 = 0 = 0\otimes V$ as well as $V\otimes F \cong V \cong F\otimes V$, i.e. F is the neutral element for the tensor product. The proof consists of showing that $\{v_iw_j : i,j\}$ is a base of $V\otimes W$ if $\{v_i : i\}$ is a base of V and $\{w_j : j\}$ is a base of W. Then it is also clear how to define the second map, namely for $f_i\colon V_i \to W_i$ we set $(f_1\otimes f_2)\colon V_1\otimes V_2 \to W_1\otimes W_2 : v_1\otimes v_2 \mapsto f_1(v_1)\otimes f_2(v_2)$ and extend F-linear to all elements of $V_1\otimes V_2$.

2. The whole construction also works for abelian groups if we replace F with $\mathbb{Z}$. The tensor product of finitely generated abelian groups is again finitely generated, however there is no notion of dimension for arbitrary abelian groups.

Proposition 4.3.3. *The tensor product is associative and commutative, i.e. for $A,B,C \in \mathfrak{Obj}$ there are natural transformations $\tau_{A,B,C}\colon ((A\otimes B)\otimes C) \xrightarrow{\sim} (A\otimes(B\otimes C))$ and $\sigma_{A,B}\colon A\otimes B \xrightarrow{\sim} B\otimes A$.*

Given one object $A \in \mathfrak{Obj}$, then $\mathbb{T}(A) := \bigoplus_{n\geq 0} A^{\otimes n}$ is a semigroup with the convention $A^{\otimes 1} := A$, $A^{\otimes 0} := F$ (or $\mathbb{Z}$ for groups).

This follows from the definition of tensor product together with the properties of an abelian category.

Example 4.3.4. Coming back to the category $\mathfrak{Vect}_F$. Then this is a monoidal category under the operations $\oplus$, direct sum – the coproduct, and $\otimes$, with neutral element F. The object $\mathbb{T}(V) := \bigoplus_{n\geq 0} V^{\otimes n}$ with $V^{\otimes 0} := F$ is an F-algebra generated by $\imath\colon V \xrightarrow{\sim} V^{\otimes 1} \subset \mathbb{T}(V)$. Unfortunately it is infinite dimensional, except when $V = 0$.

Sometimes it is also useful to consider symmetric tensor products, defined as follows:

Definition 4.3.5. *Given a vector space V/F with $2 \neq 0 \in F$, then the (anti)-symmetric tensor product is $S^2V := V \otimes V/K(V)$ ($\wedge^2 V := V \otimes V/S^2V$) where*

$$K(V) := \langle v_1 \otimes v_2 - v_2 \otimes v_1 : v_{1/2} \in V\rangle_F$$
$$S^2V \cong \langle v \otimes v : v \in V\rangle_F \subset V \otimes V.$$

Correspondingly we define the higher powers as $S^nV := V^{\otimes n}/K_n(V)$ ($\wedge^n V := V^{\otimes n}/P_n(V)$) with

$$K_n(V) := \langle v_1 \otimes \cdots \otimes (v_k \otimes v_{k+1} - v_{k+1} \otimes v_k) \otimes \cdots \otimes v_n : v_i \in V, 1 \leq k \leq n-1\rangle_F,$$
$$P_n(V) := \langle v_1 \otimes \cdots \otimes (v_k \otimes v_k) \otimes \cdots \otimes v_n : v_i \in V, 1 \leq k \leq n\rangle_F$$

and in total the (graded) symmetric algebra $S^\bullet V := \bigoplus_{n\geq 0} S^nV$ with the convention $S^0V := F$ and $S^1V := V$, as well as $\bigwedge^\bullet V := \bigoplus_{n\geq 0} \bigwedge^n V$ with $\bigwedge^0 V := F$ and $\bigwedge^1 V := V$.

Note that in particular $V \otimes V \cong S^2V \oplus \wedge^2 V$. Both constructions are associative F-algebras. $S^\bullet V \cong \mathrm{Pol}(V)$ is abelian (i.e. commutative) while $\bigwedge^\bullet V$ is graded commutative, i.e. $V_2 \wedge V_1 = (-1)^{|V_1|\,|V_2|} V_1 \wedge V_2$ for $V_i \in \bigwedge^{|V_i|} V$.

4.3.99 Exercises

Exercise 4.3.1 ($\bigwedge^\bullet V$)**.** Given a vector space V/F where F is of characteristic 0. Show that $\dim_F \bigwedge^k V = \binom{\dim_F V}{k}$, and thus $\dim_F \bigwedge^\bullet V = 2^{\dim_F V}$ in particular finite if V is finite dimensional.

Exercise 4.3.2 (S^nV)**.** Given again a vector space V/F where $2 \neq 0 \in F$. Show that $\dim_F S^k V = \binom{\dim_F V+k-1}{k}$.

4.4 Dual modules

Remember the definition of linear functional, i.e. given a vector space V/F, then any F-linear map $\alpha\colon V \to F$ is called a linear functional. Its vector space is denoted V^* and for $\dim_F V < \infty$, we can show $V^* \cong V$. This has the following consequence:

Proposition 4.4.1 (Frobenius[2]). *Given vector spaces $V, W/F$, then $\mathrm{Hom}_F(V,W) \cong V^* \otimes_F W$ canonically.*

Proof. Let $\alpha \in V^*$ and $w \in W$. We define the linear rank-1 map $f_{\alpha,w}\colon V \to W : v \mapsto \alpha(v)w$. This gives us a homomorphism $f_\bullet\colon V^* \otimes W \to \mathrm{Hom}(V,W)$ which is injective. General considerations in linear algebra show $\dim \mathrm{Hom}_F(V,W) = (\dim_F V)(\dim_F W) = (\dim_F V^*)(\dim_F W)$ thus the above map $f_\bullet$ is also surjective. This completes the proof. □

This has the following generalization to tensor categories (i.e. an additive category with a tensor product $\otimes$):

Definition 4.4.2. *Given a tensor category and an object $A \in \mathfrak{Obj}$. we say that $A^* \in \mathfrak{Obj}$ is a dual of A if for every object $B \in \mathfrak{Obj}$, we have $\mathrm{Hom}(A,B) \cong A^* \otimes B$.*

Note that once the object A^* exists it is unique up to isomorphism.

Example 4.4.3. Given finitely generated abelian groups $\mathfrak{Ab}$ with the tensor product, then the dual of an abelian group A is $A^* \cong \mathrm{Hom}(A,\mathbb{Z})$. Due to the compatibility of Hom with $\oplus$ and $\otimes$, we obtain:

$$\begin{aligned}(A \oplus B)^* &= A^* \oplus B^*,\\ (A \otimes B)^* &\cong A^* \otimes B^*,\\ C_n^* &\cong C_n\end{aligned}$$

Note however that the last two isomorphisms are in general not unique/ canonical.

[2]F.G. Frobenius *1849/10 in Berlin/Germany, †1917/8

4.5 Flat modules

4.6 Completions

4.7 Homomorphisms

4.8 Adjoint functors

Remember the definition of adjoint of a vector space homomorphism of Euclidean/ Hilbert spaces. Say $f\colon V \to W$ linear, where $(V, (.,.)_V)$ is a Euclidean space (respectively $(W, (.,.)_W)$), then $f^*\colon W \to V$ is the unique linear map with $(f^*w, v)_V = (w, fv)_W$ for all $v \in V$ and $w \in W$.

This can be generalized to functors as follows:

Definition 4.8.1. *Given two categories $\mathcal{C}$ and $\mathcal{D}$ we say that a pair of functors $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$ is adjoint if $\forall C \in \mathfrak{Obj}_{\mathcal{C}}$ and $D \in \mathfrak{Obj}_{\mathcal{D}}$*

$$\mathrm{Hom}_{\mathcal{D}}(F(C), D) \cong \mathrm{Hom}_{\mathcal{C}}(C, G(D)).$$

More specifically, we call F the left-adjoint of G and G the right-adjoint of F.

Example 4.8.2. Consider the categories $\mathfrak{Set}$ and $\mathfrak{Gp}$ with the forgetful functor $R\colon \mathfrak{Gp} \to \mathfrak{Set}$ that sends a group to its underlying set and a group homomorphism to its underlying map. Then we can define its (left)-adjoint $F\colon \mathfrak{Set} \to \mathfrak{Gp} : S \mapsto F(S)$, with

$$\mathrm{Hom}(F(S), G) \cong \mathrm{Map}(S, R(G)) \forall G \in \mathfrak{Gp}.$$

Obviously $F(S)$ is a group generated by S that does not fulfill any relations, i.e. $F(S)$ is the free group in the generators S.

4.8.99 Exercises

Exercise 4.8.1. Consider the categories $\mathfrak{Set}$ and $\mathfrak{Vect}_F$ of sets and vector spaces over a fixed field F, respectively. let again $R\colon \mathfrak{Vect}_F \to \mathfrak{Set}$ be the forgetful functor that sends a vector space to its underlying set of vectors. Express the left-adjoint $L\colon \mathfrak{Set} \to \mathfrak{Vect}_F$ in terms of elementary constructions.

4.9 Triples

Part II

Second semester

Chapter 5

Modules (5 weeks)

Chapter 6

Homological algebra (3 weeks)

6.1 Exact sequences

6.2 Pullbacks and pushouts

6.3 Projective modules

6.4 Injective modules

6.5 The injective hull

6.6 Hereditary rings

6.7 Complexes and homology

6.8 Resolutions

6.9 Interlude: Derived functors

6.9.1 Ext

6.9.2 Tor

6.10 Universal coefficient theorem

6.11 Cohomology of discrete groups

6.12 Projective dimension

6.13 Global dimension

Appendix A

Brief review of linear algebra

This is just a brief summary of some useful results of linear algebra. The proofs, further details, and examples can be found in any textbook about linear algebra.

A.1 Linear maps and dual spaces

Given two vector spaces V and W over the same field F, we denote $\mathrm{Hom}(V, W)$ the *linear maps* from V to W, i.e. all $\phi\colon V \to W$ such that

$$\phi(\lambda v + v') = \lambda\phi(v) + \phi(v') \quad \forall v, v' \in V, \lambda \in F.$$

We denote in particular the linear maps $\phi\colon V \to F$ as *linear functionals* and $V^* := \mathrm{Hom}(V, F)$ the *dual space* of V.

Proposition A.1.1. *Given a finite dimensional vector space V and a base $\{e_i\}$. Then there is a dual base $\{e^i\}$ such that $\langle e^i, e_j\rangle = \delta^i_j$ the* Kronecker symbol *which is 1 for $i = j$ and 0 otherwise.*

In particular the dual space has the same dimension as the original space.

Given a linear map $A\colon V \to W$, then its *dual / transpose* is the linear map $A^T\colon W^* \to V^*$ with

$$\langle A^T\phi, v\rangle := \langle \phi, Av\rangle$$

forall $\phi \in W^*$ and $v \in V$.

Note that the transpose of a composition

$$U \xrightarrow{B} V \xrightarrow{A} W$$

is

$$U^* \xleftarrow{B^T} V^* \xleftarrow{A^T} W^*,$$

i.e. dualization inverts the direction of mapping and has $(AB)^T = B^T A^T$.

Another observation is that when you repeat dualization, you obtain:

Corollary A.1.2 (bi-dual space)**.** *Given a vector space V, then there is a linear map $\delta\colon V \to V^{**} : v \mapsto \delta_v$ with $\langle \delta_v, \phi\rangle := \langle \phi, v\rangle$ for every $v \in V$ and $\phi \in V^*$.*

If V is finite dimensional, then δ is an isomorphism.

Note that for arbitrary topological vector spaces (where we require the linear functionals also to be continuous) we have two independent conditions, δ injective (i.e. V^* *separates* elements of V) and δ surjective. Spaces with $V^{**} = V$ are called *reflexive*.

A.2 Rank, Determinant, and invertible endomorphisms

Given a linear map $A\colon V \to W$, we denote $\operatorname{im} A := A(V) \subset W$ the *image* of A. We call $\operatorname{rk} A := \dim \operatorname{im} A$ the (row) *rank* of A. Note that $\operatorname{rk} A \leq \dim V, \dim W$. Considering the dual map $A^T\colon W^* \to V^*$, we observe

Proposition A.2.1. $\operatorname{rk} A^T = \operatorname{rk} A$.

I.e. the row rank equals the column rank.

For an endomorphism $A \in \operatorname{End}(V)$ (i.e. $W = V$) we can measure the transformation of a *volume element* $\bigwedge^n V^* = F\mathrm{vol}$ where $n := \dim V$ and vol any non-zero vector in $\bigwedge^n V^*$ and define $A^*(\mathrm{vol}) = (\det A)\mathrm{vol}$. In this way the *determinant* of an endomorphism is defined base independent. It is immediately clear that

$$\begin{aligned} \det \mathbb{1} &= 1, \\ \det \operatorname{diag}(1, \dots, 1, \lambda, 1, \dots, 1) &= \lambda, \\ \det(AB) &= (\det A)(\det B), , \\ \det(\lambda A) &= \lambda^n \det A \end{aligned}$$

for all $\lambda \in F$. Moreover an endomorphism has full rank iff its determinant does not vanish. Somewhat lengthy considerations lead to the following explicit formulas:

Proposition A.2.2. *Given the matrix representation of an endomorphism A with respect to any base of V, then its determinant computes as*

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn} \sigma \prod_{i=1}^{n} a_{i,\sigma i}.$$

In particular the determinant is linear in every row (and every column) and changes sign under exchange of two rows (columns). Moreover the determinant is the same for the dual map $\det A^T = \det A$.

Corollary A.2.3. *There are* adjunct elements (A_{ij}) *(polynomial in the matrix entries) such that for every fixed lines* i, k

$$\delta_{ik} \det A = \sum_{j=1}^{n} a_{ij} A_{kj}.$$

The corresponding formula holds for columns.

Idea of proof. Note that up to sign the adjunct elements compute as the sub-determinants

$$A_{ij} = (-1)^{i+j} \det(a_{kl})_{k \neq i, l \neq j}$$

□

Corollary A.2.4. *An endomorphism* A *is invertible iff* $\det A \neq 0$ *and in general* $A \operatorname{cod} A = (\operatorname{cod} A)A = (\det A)\mathbb{1}$ *for the matrix* $\operatorname{cod} A$ *the transpose of the above adjunct elements.*

Therefore the invertible endomorphisms form a group $\mathrm{GL}(V) := \{A \in \mathrm{End}(V) : \det A \neq 0\}$, the *general linear group*. Given subgroups $S \subset F^*$ of the multiplicative group, it is also possible to define subgroups $\det^{-1}(S) \subset \mathrm{GL}(V)$, e.g. $\mathrm{SL}(V) := \ker \det = \{A \in \mathrm{GL}(V) : \det A = 1\}$.

Beside the determinant, we can also define the *characteristic polynomial* of an endomorphism $\mathrm{ch}_A(\lambda) := \det(\lambda \mathbb{1} - A) = \lambda^n - (\operatorname{tr} A)\lambda^{n-1} \pm \cdots + (-1)^n \det A$. The operator $\operatorname{tr}\colon \mathrm{End}(A) \to F$, called the *trace*, is a linear map, i.e.

$$\operatorname{tr}(\lambda A + B) = \lambda \operatorname{tr} A + \operatorname{tr} B \quad \forall A, B \in \mathrm{End}(V), \lambda \in F,$$

and computes (base independent) as $\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$.

Corollary A.2.5. *The trace is invariant under cyclic permutations, i.e. for all* $A, B \in \mathrm{End}(V)$ *we have*

$$\operatorname{tr}(BA) = \operatorname{tr}(AB),$$

and "more" generally

$$\operatorname{tr}(A_1 \dots A_k) = \operatorname{tr}(A_2 \dots A_k A_1).$$

Note that the characteristic polynomial ch_A for an endomorphism $A \in \mathrm{End}(V)$ also has the following property

$$\mathrm{ch}_A(A) = 0.$$

But it is not the only polynomial with that property. Obvious other polynomials would be $\mathrm{ch}_A \cdot F[x]$, but as the example $A = \mathbb{1}$ shows there may be even smaller nonzero polynomials with that property. Namely we define the *minimal polynomial* of an endomorphism $A \in \mathrm{End}(V)$ over a field F as a non-zero polynomial $p_A \in F[x]$ with $p_A(A) = 0$ and such that the degree of p_A is minimal. We can make the minimal polynomial unique by requiring it to be monic (leading coefficient 1).

Note that the roots $\lambda \in \bar{F}$ of the characteristic polynomial ch_A are the eigenvalues of A. Each eigenvalue has at least one eigenvector $v \in \bar{V} \cong \bar{F}^n$, i.e. a solution of $Av = \lambda v$. But then also $A^k v = \lambda^k v$ and thus $p_A(\lambda)v = p_A(A)v = 0$, i.e. every eigenvalue is already a root of the minimal polynomial.

Example A.2.6. In terms of the block decomposition of the Jordan normal form of A, we need a factor $(x - \lambda)^{n'}$ for every eigenvalue $\lambda \in \bar{F}$ of A and $1 \leq n' \leq n$ is the largest size of a Jordan block with eigenvalue λ. If F is not algebraically closed, we replace $(x - \lambda)$ by the minimal polynomial of λ.

A.3 Euclidean vector spaces and Hilbert spaces

A *Euclidean* vector space (*Hilbert* space) is a vector space over a field F (say real or complex numbers) together with a (conjugate) *symmetric positive-definite* (*sesquilinear*) *bilinear* form $g\colon V \times V \to F$. The conditions read for $v, w, w' \in V$ and $\lambda \in F$

$$g(w, v) = \overline{g(v, w)}, \tag{A.1}$$

$$g(v, \lambda w + w') = \lambda g(v, w) + g(v', w) \tag{A.2}$$

$$g(v, v) \geq 0, \quad \text{and “}= 0\text{” iff } v = 0. \tag{A.3}$$

Example A.3.1.

1. Given any base $\{e_i\}$ of V, we can define the standard inner product as $g(e_i, e_j) := \delta_{ij}$ with the Kronecker symbol. Now g extends uniquely to arbitrary vectors v and w using the bilinearity.

2. Given one inner product g as above and a symmetric $n \times n$-matrix B, then $b(v, w) := g(v, Bw)$ is another symmetric bilinearform. If B is positive-definite, then b is also positive-definite.

The Theorem of Gram–Schmidt states that we can find adapted bases for the inner product as follows:

Theorem A.3.2 (Gram–Schmidt)**.** *Given a Euclidean vector (Hilbert) space (V, g) and a countable generating system for V, then we can reduce this to an orthonormal (unitary) base $\{e_i\}$, i.e. $g(e_i, e_j) = \delta_{ij}$.*

It is also possible to generalize the notions to *indefinite* non-degenerate symmetric bilinear-forms as follows. Instead of (A.3), we require the weaker condition

$$g^{\#}\colon V \xrightarrow{\sim} V^*, (g^{\#}v)(w) := g(v,w).$$

Note that the isomorphism property for finite dimensional (reflexive) vector spaces is equivalent to $g^{\#}v = 0$ iff $v = 0$, i.e. non-degeneracy of $g^{\#}$.

A.3.1 Adjoint map and Orthogonal/ Unitary transformations

With respect to any such (non-degenerate) inner product, we define the *adjoint* of a linear map $A\colon V \to V$ as $A^\dagger\colon V \to V$ with

$$g(v, A^\dagger w) := g(Av, w) \quad \forall v, w \in V.$$

Note that analogously to the transpose map $(AB)^\dagger = B^\dagger A^\dagger$ and $(A+B)^\dagger = A^\dagger + B^\dagger$. For the complex numbers however the map $\dagger\colon \mathrm{End}(V) \to \mathrm{End}(V)$ is not $\mathbb{C}$-linear, but antilinear, i.e. $(\lambda A)^\dagger = \bar{\lambda} A^\dagger$. In the case of a symmetric bilinear form the adjoint is the same as the transpose. This is mathematically rigorous, because for a Euclidean vector space $V^* \cong V$, canonically (via $g^{\#}$).

This gives us further means to study symmetric bilinear forms on vector spaces.

Proposition A.3.3. *Given a Euclidian (Hermitean) vector space (V, g), then any symmetric bilinear (sesquilinear) form b can be encoded as $b(v,w) := g(v, Bw)$ where $B \in \mathrm{End}(V)$ is any symmetric endomorphism $B^\dagger = B$. The change of the matrix associated to B under a coordinate change $A \in \mathrm{GL}(V)$ is $A^*(B) = A^\dagger B A$.*

We can conversely also ask for those automorphisms $A \in \mathrm{GL}(V)$ that leave the symmetric bilinear form invariant and thus define $\mathrm{O}(V,g) := \{R \in \mathrm{GL}(V) : R^* g = g\}$ the *Orthogonal group*. Noting that w.r.t. any orthonormal basis the inner product of a Euclidean vector space can be encoded via the identity matrix $B = \mathbb{1}$, we see that the orthogonality relation is (canonically isomorphic to) $R^T R = \mathbb{1}$. Note that for any finite dimensional vector space V starting with any $R \in \mathrm{End}(V)$ this condition assures that V is bijective, hence invertible, i.e. $R \in \mathrm{GL}(V)$. But then also $RR^T = \mathbb{1}$ follows.

We define $\mathrm{O}(n) := \mathrm{O}(F^n, \langle .,. \rangle)$ the orthogonal group on the standard vector space F^n w.r.t. the standard inner product. But there are also the $\mathrm{O}(p,q) := \mathrm{O}(F^{p+q}, \eta)$ where $\eta = \mathrm{diag}(1, \ldots, 1, -1, \ldots, -1)$ is the (generalized) Minkowski metric with p positive entries and q negative entries.

The condition $A^T A = \mathbb{1}$ implies $(\det A)^2 = 1$ and therefore $\det \mathrm{O}(V) = \{\pm 1\}$ (for $\dim V \geq 1$). The intersection $\mathrm{SO}(V) := \mathrm{O}(V) \cap \mathrm{SL}(V)$ reduces thus to half of the elements, the *orientation preserving* orthogonal transformations.

Analogously for Hermitean vector spaces, we can ask for the coordinate changes $A \in \mathrm{GL}(V)$ that leave the Hermitean structure h invariant, i.e. $A^* h = h$ – which we denote as *unitary transformations* $\mathbb{U}(V,h)$. Again w.r.t. any unitary basis, the Hermitean form can be encoded via $\mathbb{1}$ and the unitary condition is thus $A^\dagger A = \mathbb{1}$. Also for finite dimensional complex vector spaces any $A \in \mathrm{End}(V)$ fulfilling the condition is invertible and thus also $AA^\dagger = \mathbb{1}$.

Correspondingly, we can define the $\mathbb{U}(n) := \mathbb{U}(\mathbb{C}^n, \langle .,. \rangle)$ the unitary transformations of the standard complex vector space $\mathbb{C}^n$ w.r.t. the standard Hermitean structure. Moreover there are the $\mathbb{U}(p,q) := \mathbb{U}(\mathbb{C}^{p+q}, \eta_{\mathbb{C}})$ where $\eta_{\mathbb{C}}$ is the pseudo-Hermitean extension of the (generalized) Minkowski metric η.

The condition $A^\dagger A = \mathbb{1}$ implies $|\det A|^2 = 1$ and thus $\det \mathbb{U}(V) = \mathbb{S}^1 = \mathbb{U}(1)$ (for $dim V \geq 1$). The intersection $\mathrm{SU}(V) := \mathbb{U}(V) \cap \mathrm{SL}(V)$ reduces thus the dimension by 1, the volume preserving unitary transformations.

Theorem A.3.4 (Sylvester)**.** *Given a (finite dimensional) vector space V over the real numbers, then its symmetric bilinear forms $b \in S^2 V^*$ are characterized by its defect index $d := \dim \ker b^\#$ and its signature $\mathrm{sgn}\, b := n_+ - n_-$ where $n_\pm$ is the number of positive/ negative unit vectors in the normal form*

$$b = e_1^2 + \cdots + e_{n_+}^2 - f_1^2 - \cdots - f_{n_-}^2$$

with $\dim V = n_+ + n_- + d$.

The same formulas are true for pseudo-Hermitean forms.

For complex Euclidean vector spaces however, there is no signature, because we can change every negatively normed vector into a positively normed one.

A.3.2 Isometries

We can also ask the question for all *isometries* of $E^n := (\mathbb{R}^n, \langle .,. \rangle)$. Here we understand E^n as an *affine space* (i.e. forget about the special role of the origin) and have $\mathbb{R}^n$ acting transitively on it. The inner product $\langle .,. \rangle$ now measures tangent vectors $v, w \in T_x E^n \cong \mathbb{R}^n$. Given any fixed point $0 \in E^n$, we see that the isometries preserving 0 are $\mathrm{ISO}_0(n) \cong \mathrm{O}(n)$ which break up into *rotations* $\mathrm{SO}(n)$ and *reflections* $\det^{-1}(-1) = \mathrm{O}(n) \setminus \mathrm{SO}(n)$.

But there are also *translations* which are uniquely characterized by a vector in $\mathbb{R}^n$. Note that the orthogonal transformations act on the group of translations, i.e. in particular for every $R \in \mathrm{O}(n)$, $R\mathbb{R}^n R^{-1} \subset \mathbb{R}^n$ and thus the isometries are a group extension:

$$0 \to \mathbb{R}^n \overset{\lhd}{\hookrightarrow} \mathrm{ISO}(n) \twoheadrightarrow \mathrm{O}(n) \to 1.$$

Note however that rotations/ reflections do not commute with all translations. Therefore the product is not a direct product, but rather a semi-direct product

$\mathrm{ISO}(n) = \mathrm{O}(n) \ltimes \mathbb{R}^n$. The group law is $(R, v)(R', v') = (RR', v + Rv')$, the neutral element $\mathrm{Id} = (\mathbb{1}, 0)$ and the inverses are $(R, v)^{-1} = (R^{-1}, -R^{-1}v)$.

The corresponding notions for the Minkowski space $M^{1,d} = (\mathbb{R}^{1+d}, \eta)$ are $\mathrm{ISO}(1, d) \cong \mathrm{O}(1, d) \ltimes \mathbb{R}^{1+d}$ the *Poincaré group* and $\mathrm{ISO}_0(1, d) \cong \mathrm{O}(1, d)$ the *Lorentz transformations.*

A.4 Symplectic vector spaces

Beside symmetric bilinear forms, we can also consider a skew-symmetric non-degenerate bilinear form $\omega\colon \bigwedge^2 V \to F$, i.e. we require

$$\omega(w, v) = -\omega(v, w), \tag{A.4}$$

$$\omega(v, \lambda w + w') = \lambda\omega(v, w) + \omega(v, w') \tag{A.5}$$

$$\omega^\#\colon V \xrightarrow{\sim} V^*, (\omega^\# v)(w) := \omega(v, w) \tag{A.6}$$

Note that $\omega(v, v) = 0$. However the isomorphism property assures the non-degeneracy condition $\omega^\#(v) = 0$ iff $v = 0$. Therefore analogously to the Euclidean vector spaces, the symplectic vector spaces are self-dual (via $\omega^\#$).

The analogue of Sylvester's Theorem is the following observation:

Proposition A.4.1 (Darboux)**.** *Given a symplectic vector space (V, ω). Then there is a base $\{p_1, \ldots, p_n, v^1, \ldots, v^n\}$ such that*

$$\omega(p_i, v^j) = \delta_i^j, \quad \omega(p_i, p_j) = 0 = \omega(v^i, v^j).$$

In particular the dimension of every symplectic vector space is even.

We can also ask for the automorphisms of V that preserve the symplectic structure ω and end up with $\mathrm{SP}(V, \omega) := \{A \in \mathrm{GL}(V) : A^*(\omega) = \omega\}$. Given the above Darboux coordinates, we see that $\omega(v, w) = \langle v, Jw\rangle$ w.r.t. the symmetric inner product $\langle ., .\rangle$ and the matrix $J := \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}$. The condition of preserving the symplectic form thus reads $A^T JA = J$. We see that the determinant of J is $\det J = 1$ and thus for any $A \in \mathrm{End}(V)$ that is symplectic then $(\det A)^2 = 1$, i.e. in particular A is invertible.

Proposition A.4.2 (Liouville)**.** $\det \mathrm{SP}(V, \omega) = 1$ *for real (complex) symplectic vector spaces.*

The corresponding symplectomorphisms of the affine symplectic space $(\mathbb{A}^{2n}, \omega)$ are $\mathrm{SP}(n) \ltimes F^{2n}$ where F is the base field (real or complex numbers).

Appendix B

Zorn's lemma in algebra

Remember that we are still missing the proof of the existence of a field extension $F \subset E$ that contains all algebraic elements over F. As the Lemma 3.2.7 shows, for every particular algebraic element α over F there is (a minimal polynomial and thus) an algebraic field extension $E_\alpha := F[\alpha]/(p)$ that contains a root of the minimal polynomial and thus $\alpha \in E_\alpha$ unique up to isomorphism.

The idea is now to add more and more algebraic elements and to take the limit in some sense. The precise notion of this limit is captured by the following Property:

Definition B.0.1 (Zorn's[1] Lemma)**.** *Given a nonempty partially ordered set* $(X, \leq)$ *and every non-empty chain* $x_1 \leq x_2 \leq \ldots$ *of elements* $x_i \in X$ *has an upper bound* $u \in X$*, i.e. all* $x_i \leq u$*, then* X *has a maximal element* $M \in X$*, i.e. for all* $x \in X$*,* $x \leq M$*.*

Every finite set with every partial order fulfills Zorn's lemma, as can be proved by induction. For infinite sets it is less intuitive. In fact it is independent to the other axioms of set theory, except for:

Proposition B.0.2. *Given the axioms of set theory, then Zorn's lemma is equivalent to the axiom of choice.*

Definition B.0.3 (Axiom of choice)**.** *Given a small family* $\mathcal{F}$ *of non-empty sets, i.e. for all* $S \in \mathcal{F}$ *there is an element* $x \in S$*. Then there is a choice function* $c\colon \mathcal{F} \to \bigcup_{S\in\mathcal{F}} S$ *such that for every* $S \in \mathcal{F}$*,* $c(S) \in S$*.*

This condition may be a bit more intuitive, because it seems "evident" that one can choose an element from each non-empty set.

We will however omit the proof of the equivalence of the two axioms.

[1] Max A. Zorn *6/1906 in Krefeld/ Germany, †3/1993

Instead we want to show how Zorn's lemma leads to a proof of the existence of the algebraic closure of a field F (Theorem 3.2.8). Referring to the partial proof in Section 3.2 it remains to show the following:

Lemma B.0.4. *Given a field F, there is a field $E \supset F$ that has for every non-constant polynomial $p \in F[x]$ a root $\alpha \in E$.*

Proof. Start from the category X_0 of all finite algebraic extensions of F, i.e. for every irreducible polynomial $p \in F[x]$ there is a field $E_p := F[\xi]/(p)$ that has a root ξ of p. If we identify the isomorphic extensions, we end up with a set (small category) X_1, because the irreducible polynomials over F are a subset of $F[x]$ and each of these polynomials has a finite set of roots. The partial order of X_0 carries over to X_1 and the finite chains in X_1 do have maximal elements, the last element. However the infinite chains do not yet have an upper bound, but we can create them by defining $X := \{\bigcup_{n\geq 0} E_n : E_0 := F \subset E_1 \subset E_2 \subset \cdots, \forall n : E_n \in X_1\}$. The partial order carries over from X_1 to X and now all chains in X have an upper bound.

Now by Zorn's lemma there is a maximal element $E \in X$ such that for every $K \in X$, $F \subset K \subset E$. But then in particular also every polynomial $p \in F[x]$ has a root in some $E_p \in X_1 \subset X$ and since $E_p \subset E$, this E has the desired properties. □

Another point of omission was the proof of Proposition 3.2.4 for infinite algebraic field extensions. This can be done in a similar fashion as follows:

Proof of Proposition 3.2.4. As shown in Section 3.1 every finite field extension $K \subset E' \subset E$ permits an extension ϕ' of ϕ to E'. Let thus X_1 be the set of all such extensions. It has a partial order via $\phi_1 \leq \phi_2$ iff $D(\phi_1) \subset D(\phi_2)$ and again the finite non-empty chains in X_1 have an upper bound. In order to also obtain upper bounds for infinite chains, consider the completion $X := \{(\bar{\phi}\colon E_\infty \to E) : \phi = \phi_0 \leq \phi_1 \leq \phi_2 \leq \ldots, \phi_n\colon E_1 \to E, E_\infty := \bigcup_{n\geq 0} E_n, \bar{\phi}(a) = \phi_n(a) \text{ for } a \in E_n\}$. Now every chain has an upper bound $\bar{\phi}$ and so by Zorn's lemma there is a maximal element $\sigma \in X$ with $\sigma\colon E_\infty \to E$ such that for every $\tilde{\phi}\colon E' \to E$ and $k'\colon E' \hookrightarrow E$, $\tilde{\phi} = \sigma \circ k'$. If $E_\infty \subsetneq E$ then there is another algebraic element in $E \setminus E_\infty$ which is a contradiction to σ being maximal. Since σ is a ring-endomorphism of the field E it must be an automorphism. □

Bibliography

[Gri07] P. Grillet: *Abstract algebra*, Graduate texts in mathematics, Springer (**2007**), ISBN 9780387715674.

[Kol48] E. R. Kolchin: *Algebraic matrix groups and the Picard–Vessiot theory of homogeneous linear ordinary differential equations*, Annals of Mathematics, vol. 49, pp. 1–42 (**1948**), ISSN 0003-486X.

[PS03] M. van der Put and M. Singer: *Galois theory of linear differential equations*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] (**2003**), URL `http://www4.ncsu.edu/~singer/ms_papers.html`.

[web] written by the web: *Wikipedia the free encyclopedia*, URL `http://www.wikipedia.org/`.