Introduction to Algebra

Lothar Göttsche

Contents

CHAPTER 1

Groups

In this chapter we introduce one of the most fundamental and important concepts in algebra and in the whole of mathematics: groups. A group will be a set G such that we can multiply the elements of G. That is there is an operation

$$\cdot : G \times G \to G, (a, b) \mapsto a \cdot b,$$

which satisfies some axioms. It generalizes many examples that we know from high school or linear algebra.

(1) $(\mathbb{Z}, +)$ the integers with the usual addition,
(2) $(\mathbb{Q} \setminus \{0\}, \cdot)$ the nonzero rational numbers with the usual multiplication,
(3) $n \times n$ matrices over $\mathbb{R}$ with nonzero determinant with matrix multiplication

Groups play an important role in almost all parts of mathematics, and many more complicated structures (like rings, fields) are build on groups.

1. Definition of a group

DEFINITION 1.1. A nonempty set G together with a binary operation

$$\cdot : G \times G \to G;\ (a, b) \mapsto a \cdot b$$

(the multiplication or product) is called a *group*, if the following axioms hold.

(1) (Associativity) for all $a, b, c \in G$ we have
$$(a \cdot b) \cdot c = a \cdot (b \cdot c).$$
(the left-hand-side means: multiply a and b first, then with c, similar for the right-hand-side).
(2) (Neutral element) There exists an element $e \in G$, such that $e \cdot a = a \cdot e = a$ for all $a \in G$. e is called the neutral element of e.
(3) (Inverse) For every $a \in G$ there exists an element $a^{-1} \in G$ with $a \cdot a^{-1} = a^{-1} \cdot a = e$. a^{-1} is called the inverse of a.

NOTATION 1.2. (1) In future we will often drop the symbol $\cdot$ for the multiplication, i.e. we write ab instead of $a \cdot b$.
(2) Note also that because of associativity we can drop brackets in the notation, i.e. we can write $abc := (ab)c = a(bc)$.

As an exercise in the definitions we show the uniqueness of the neutral and the inverse element.

REMARK 1.3. (1) The neutral element is unique.
(2) The inverse a^{-1} of a is unique.
(3) Let b be an element with $ba = e$, then $b = a^{-1}$.
(4) $(a^{-1})^{-1} = a$.
(5) $(ab)^{-1} = b^{-1}a^{-1}$.

PROOF. (1) Let e, e' be neutral elements of G. Then $e' = ee' = e$.
(2) Let b, b' be inverses of a. Then $b' = b'e = b'(ab) = (b'a)b = eb = b$.
(3) Let b be an element with $ba = e$, Let c be an element with $cb = e$. Then

$$ab = eab = cbab = ceb = cb = e,$$

thus $ab = ba = e$, i.e. $b = a^{-1}$.
(4) a satisfies $aa^{-1} = e$, and thus it is the inverse of a^{-1}.
(5) $abb^{-1}a^{-1} = aa^{-1} = e$. □

A useful property of groups is the cancellation property, i.e we can cancel factors on both sides of an equation in G.

PROPOSITION 1.4. *Let G be a group and $a, b, c \in G$.*
(1) *If $ab = ac$, then $b = c$.*
(2) *If $ba = ca$, then $b = c$.*

PROOF. (1) We multiply $ab = ac$ on both sides with a^{-1}. We get $b = a^{-1}ab = a^{-1}ac = c$. (2) is similar. □

I very nice property that many groups have is that the order of multiplication is not important.

DEFINITION 1.5. A group G is *commutative* (or *abelian*) if $a \cdot b = b \cdot a$ for all $a, b \in G$.

NOTATION 1.6. Often the group operation in abelian groups is written $a + b$. In this case one writes $-a$ instead of a^{-1} and $a - b$ instead of ab^{-1}.

Examples of groups

(1) The trivial group $\{1\}$ consists of one element 1 with $1 \cdot 1 = 1$.
(2) $(\mathbb{Z}, +)$, the integers with the usual addition, form a abelian group. The neutral element is 0 and the inverse of a is $-a$.
(3) The nonzero rational numbers $\mathbb{Q} \setminus \{0\}$ with the usual multiplication

$$\frac{m}{n}\frac{m'}{n'} = \frac{mm'}{nn'}$$

form an abelian group. The neutral element is $1 = \frac{1}{1}$ and the inverse of $\frac{m}{n}$ is $\frac{n}{m}$.

(4) Similarly the real numbers $(\mathbb{R}, +)$ with usual addition form a group, so does $(\mathbb{R} \setminus \{0\}, \cdot)$ the nonzero real numbers with the usual multiplication. Both are abelian groups.

(5) Let $k \in \mathbb{Z}$ and let $\mathbb{Z}_k := \{0, 1, \dots, k-1\}$. We define a commutative group structure on $\mathbb{Z}_k$ as follows. For an integer n let $\underline{n} \in \mathbb{Z}_k$ be the rest of n when divided by k, i.e. we write $n = dk + \underline{n}$ with $d \in \mathbb{Z}$ and $0 \leq \underline{n} < k$. We define an addition $\oplus$ on $\mathbb{Z}_k$ by $n \oplus m = \underline{n+m}$, where $+$ is the usual addition in $\mathbb{Z}$., We can check that this makes $\mathbb{Z}_k$ into an abelian group.
   (a) $\underline{n+m} \oplus \underline{l} = \underline{n+m+l} = \underline{n} \oplus \underline{m+l}$, thus the addition is associative.
   (b) $0 \oplus n = \underline{0+n} = \underline{n} = n$, so 0 is the neutral element.
   (c) $n \oplus (k-n) = \underline{k} = 0$. So $k-n$ is the inverse of n.
   (d) Clearly $\underline{n+m} = \underline{m+n}$, so the group is commutative.

(6) The 2×2 matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with real coefficients with $det(A) = ad - bc \neq 0$ form a group under matrix multiplication

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix}.$$

The neutral element is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and the inverse of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is

$$\begin{pmatrix} \frac{d}{ad-bc} & -\frac{c}{ad-bc} \\ -\frac{b}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}.$$

This group is not commutative:

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

(7) Let M be a set. The *group of permutations* $S(M)$ of M (or the *symmetric group on* M)is the set of bijective maps $f : M \to M$. This is a group with the composition of maps $f \cdot g := f \circ g$ as multiplication:
   (a) It is well-known that the composition of maps is always associative $(f \circ g) \circ h = f \circ (g \circ h)$.
   (b) the neutral element is the identity $id_M : M \to M; m \mapsto m$.
   (c) every $f \in S(M)$ is bijective, and therefore has an inverse map f^{-1}, and by definition $f^{-1} \circ f = f \circ f^{-1} = id_M$. Thus f^{-1} is the inverse element of f.

(8) A particularly important case is when M is the set $\{1, \dots, n\}$. Then $S_n := S(\{1, \dots, n\})$ is called the *symmetric group* of degree n. It consist of the permutations of $\{1, \dots, n\}$.

We introduce the following notation for permutations. If $f : \{1, \dots, n\} \to \{1, \dots, n\} \in S_n$ is a bijection, we write it as

$$\begin{pmatrix} 1 & 2 & \dots & n \\ f(1) & f(2) & \dots & f(n) \end{pmatrix}.$$

Thus by definition

$$e = \begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix}.$$

is the neutral element of S_n.

In particular we see that S_3 consists of the elements

$$e = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

We see that S_3 is not commutative:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Groups were originally introduced as groups of symmetry and were usually groups of permutations of some set.

(9) Let G,H be groups. The product $G \times H$ consists of the pairs g, h, with $g \in G, h \in H$. It is a group. The group operation is componentwise

$$(g, h) \cdot (g', h') = (gg', hh').$$

It is easy to check that with this operation $G \times H$ is a group. The neutral element is (e, e'), where e is the neutral element of G and e' is the neutral element of H . The inverse of (g, h) is (g^{-1}, h^{-1}).

An important invariant of a group is the number of elements it contains.

Definition 1.7. Let G be a group. The *order* of $ord(G)$ of G is the number of elements of G, we write $ord(G) = \infty$, if G contains infinitely many elements. Otherwise we call G a *finite group*.

Exercise 1.8. Show $ord(S_n) = n!$.

Notation 1.9. We introduce a shorthand for powers of elements in a group. Let G be a group and $a \in G$. We define a^n for $n \in \mathbb{Z}$ any integer. If $n > 0$ we write

$$a^n = \underbrace{a \cdot a \cdot \ldots \cdot a}_{n \text{ times}}.$$

We put $a^0 = e$ and $a^{-n} = (a^{-1})^n$.

If G is an abelian group and we write the group operation as $a + b$, then we write na instead of a^n for $n \in \mathbb{Z}$.

EXERCISE 1.10. Show that $a^n \cdot a^m = a^{n+m}$ for all $n, m \in \mathbb{Z}$, and $a^{-n} = (a^n)^{-1}$ for $n \in \mathbb{Z}$.

2. Subgroups

NOTATION 2.1. From now on we usually write 1 for the neutral element of a group. If the group is commutative and we write the group operation as $+$ we usually write 0 for the neutral element.

We want to consider subsets H of a group G, but only those which somehow reflect the fact that G is a group. The most natural condition is that H with the restriction of the multiplication on G itself forms a group.

DEFINITION 2.2. A nonempty subset H of a group G is called a *subgroup* of G, if for any two elements $a, b \in H$ we have $a \cdot b \in H$, and H with the restriction of $\cdot$ forms a group.

REMARK 2.3. Obviously if $H \subset G$ is a subgroup and $L \subset H$ is a subgroup, then $L \subset H$ is a subgroup.

We need a criterion for a subset $H \subset G$ to be a subgroup.

LEMMA 2.4. *A nonempty subset H of a group G is a subgroup of G if and only if*

(1) *if $a, b \in H$ then $ab \in H$,*
(2) *if $a \in H$, then $a^{-1} \in H$.*

PROOF. If H is a subgroup of G, then (1) and (2) must obviously hold.

Conversely let $H \subset G$ be a subset satisfying (1) and (2). The associative law holds for all elements $a, b, c \in G$, in particular it holds in H. Given a in H we have $a^{-1} \in H$ and thus $1 = aa^{-1} \in H$, so H has a neutral element and a has an inverse in H. □

Examples of subgroups

(1) Let $(\mathbb{Z}, +)$ be the group of integers with addition. Let $k \in \mathbb{Z}$ be an integer. Then
$$k\mathbb{Z} := \{kn \mid n \in \mathbb{Z}\}$$
is the set of integers divisible by k. Obviously if n and m are divisible by k, so is ab and so is $-a$. Thus $k\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.
(2) $(\mathbb{R}_{>0}, \cdot)$ is a subgroup of $(\mathbb{R}^*, \cdot)$.
(3) $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{R}, +)$.
(4) Let H_1, H_2 be subgroups of a group G, then $H_1 \cap H_2$ is a subgroup of G (exercise).

(5) Let G be a group, and let $a \in G$ be an element. Let

$$\langle a \rangle := \{a^n \mid n \in \mathbb{Z}\}.$$

By $a^n \cdot a^m = a^{n+m}$, $(a^n)a^{-n} = 1$, we see that $\langle a \rangle$ is a subgroup of G. It is also easy to see that $\langle a \rangle$ is abelian.

$\langle a \rangle$ is called the *cyclic subgroup* of G generated by $\langle a \rangle$.

For instance $k\mathbb{Z} \subset \mathbb{Z}$ is the cyclic subgroup generated by k.

(exercise).

DEFINITION 2.5. A group G is called *cyclic*, if there is an element $a \in G$ with $G = \langle a \rangle$.

EXAMPLE 2.6. $\mathbb{Z}_k$ is cyclic: It is generated by 1.

DEFINITION 2.7. More generally any subset of a group determines a subgroup. Let G be a group, $U \subset G$ be a subset. The *subgroup of G generated by U* is the smallest subgroup of G containing U (technically it is the intersection of all subgroups containing U, it exists, because G is a subgroup containing U). We denote it by $\langle U \rangle$. If $g_1, \ldots, g_r \in G$, we call $\langle g_1, \ldots, g_r \rangle := \langle \{g_1, \ldots, g_r\} \rangle$ the subgroup of G generated by $g_1, \ldots, g_r$.

Now we want to see that a subgroup H of a group G determines a decomposition of G into disjoint smaller subsets.

First I want to review the notion of an equivalence relation.

Review of equivalence relations and equivalence classes

The concept of equivalence relation is very basic and important, and used in every field of mathematics.

DEFINITION 2.8. Let A be a set. A *relation* on A is a subset $R \subset A \times A$. We write $a \sim b$ if $(a, b) \in R$. The relation is called an *equivalence relation* if the following axioms hold:

(1) (reflexivity) $a \sim a$ for all $a \in A$,
(2) (symmetry) if $a \sim b$, then $b \sim a$.
(3) (transitivity) if $a \sim b$ and $b \sim c$, then $a \sim c$.

We say a is *equivalent to* b if $a \sim b$.

In practice when trying to show that a relation is an equivalence relation, usually reflexivity and symmetry are easy, and transitivity is more subtle to check.

EXAMPLE 2.9. Let $\mathbb{Z}$ be the set of integers, and let k be an integer. For integers a, b, we say that $a \sim b$ if $a - b$ is divisible by k. Check hat $\sim$ is an equivalence relation.

The main reason we consider equivalence relations is that an equivalence relation on a set A determines a decompostion of A into disjoint subsets, the equivalence classes.

DEFINITION 2.10. Let A be a set and let $\sim$ be an equivalence relation on A. The *equivalence class* of $a \in A$ is the set

$$[a] := \big\{ b \in A \;\big|\; a \sim b \big\}.$$

We write $A/\sim$ for the set of equivalence classes.

EXAMPLE 2.11. Let $k \in \mathbb{Z}_{>0}$ and let again $\sim$ be the equivalence relation on $\mathbb{Z}$ given by $a \sim b$, if $a - b$ is divisible by k. Let $\underline{a} \in \{0, \ldots, k-1\}$ be the rest when dividing a by k. Check that the equivalence class of $[a]$ of a is the set of all $b \in \mathbb{Z}$ with $\underline{b} = \underline{a}$. Thus

$$\mathbb{Z} = [0] \cup [1] \cup \ldots \cup [k-1].$$

is a decomposition of $\mathbb{Z}$ into disjoint equivalence classes.

This example generalizes: an equivalence relation on A determines a decomposition of A into disjoint equivalence classes.

PROPOSITION 2.12. *Let A be a nonempty set, and $\sim$ an equivalence relation on A. Then the distinct equivalence classes $[a]$ are a decomposition of A into disjoint nonempty subsets.*

PROOF. As $a \sim a$, we see that $a \in [a]$, so clearly the union of all equivalence classes is A. We only have to check that for $a, b \in A$, either $[a] = [b]$ or $[a] \cap [b] = \emptyset$. Thus assume $[a] \cap [b] \neq \emptyset$. So let $x \in [a] \cap [b]$. Thus we have $a \sim x$, and $b \sim x$, thus also $x \sim b$. By transitivity this gives $a \sim b$. Thus if $y \in [a]$, then $y \sim a$ and by transitivity and $a \sim b$ we get $y \sim b$, i.e. $y \in [b]$. Thus $[a] \subset [b]$. In the same way we see $[b] \subset [a]$, thus $[a] = [b]$. □

DEFINITION 2.13. Let A be a nonempty set and $\sim$ an equivalence relation on A. Let $[a]$ be the equivalence class of an element $a \in A$ (in other words $a \in [a]$). Then we say a is a *representative of* $[a]$. If $a' \in [a]$, also a' is a representative of $[a]$.

Often one wants to define something for equivalence classes $[a]$ in terms of the representatives a. This will be a valid definition, if and only if one gets the same definition for any choice of representative $a' \in [a]$, in other words, if the definition is independent of the representative.

EXAMPLE 2.14. Let k be a positive integer. On the integers $\mathbb{Z}$ we define an equivalence relation $\equiv \mod k$ by $n \equiv m \mod k$ if and only if $n - m$ is divisible by k. It is easy to see that this is an equivalence relation. We denote by $[n]$ the equivalence class of n, and by $\mathbb{Z}/k\mathbb{Z}$ the set of equivalence classes.

We want to make $\mathbb{Z}/k\mathbb{Z}$ into a group, by defining $[n] + [m] = [n+m]$. Note that we have defined the sum in terms of representatives, so the definition makes sense if and only if it is independent of the representatives, in other words if it is well-defined. So let $[n'] = [n]$, $[m'] = [m]$. This means $n' \equiv n \mod k$, $m' \equiv m \mod k$. Then we need to see $[n'+m'] = [n+m]$, i.e. $n' + m' \equiv n + m \mod k$.

This is trivially true: if $n - n' = dk$ and $m - m' = ek$ for integers d, e, then $(n + m) - (n' + m') = (d + e)k$. Thus $[n' + m'] = [n + m]$.

We can easily check that this makes $\mathbb{Z}/k\mathbb{Z}$ into a group because $\mathbb{Z}$ is:

$$([n] + [m]) + [l] = [n + m] + [l] = [n + m + l] = [n] + [m + l] = [n] + ([m] + [l]),$$
$$[0] + [m] = [m], \quad [m] + [-m] = [0].$$

Cosets

Now we want to see that a subgroup $H \subset G$ determines an equivalence relation on G and therefore a decomposition of G into equivence classes called cosets.

DEFINITION 2.15. Let G be a group and let H be a subgroup of G. For elements $a, b \in G$ we say a *is congruent to* $b \bmod H$, written $a \equiv b \bmod H$ if there is an h in H with $a = bh$.

It is very easy to see that conguence mod H is an equivalence relation on G: Let $a, b, c \in G$.

(1) (reflexivity) $a = a1$ and $1 \in H$ because H is a subgroup, thus $a \equiv a \bmod H$.
(2) (symmetry) Assume $a \equiv b \bmod H$, i.e. $a = bh \in H$. Then $b = ah^{-1}$, and $h^{-1} \in H$, thus $b \equiv a \bmod H$.
(3) (transitivity) Assume $a \equiv b \bmod H$ and $b \equiv c \bmod H$, i.e. $a = bh_1$, $b = ch_2$, with $h_1, h_2 \in H$. Then $a = ch_2h_1$ and $h_2h_1 \in H$, thus $a \equiv c \bmod H$.

Thus H determines a decomposition of G into disjoint equivalence classes mod H. The equivalence classes will be called the cosets.

DEFINITION 2.16. Let H be a subgroup of G. For every $a \in G$ the set

$$aH := \{ah \mid h \in H\}$$

is called a (left) *coset* of H in G.

By definition aH is just the equivalence class

$$[a] = \{x \in G \mid a \equiv x \bmod H\}.$$

Note that by definition $H = 1H$.

Thus by the above we have a decompostion of G into disjoint cosets aH.

REMARK 2.17. Let aH, bH be two cosets of H in G. Then the map

$$aH \to bH; ah \mapsto bh$$

is a bijection.

PROOF. Clearly the map is onto, and it is injective because if $bh_1 = bh_2$, then by the cancelation property $h_1 = h_2$, thus $ah_1 = ah_2$. □

It is now easy to prove a well-known theorem due to Lagrange: the order of a subgroup of a finite group divides the order of the group.

DEFINITION 2.18. Let H be a subgroup of a group G. The *quotient set* G/H is the set of left cosets of H. The *index* of H in G is $[G : H] = |G/H|$ the number of left cosets of H in G, where we put $[G : H] = \infty$ if G/H is infinite.

It can happen that H is an infinite subgroup of an infinite group G with $[G : H]$ finite.

EXAMPLE 2.19. Let k be an integer. $k\mathbb{Z}$ is a subgroup of $\mathbb{Z}$. The equivalence relation in Example **??** is just $n \equiv m \mod k$. Thus the set $\mathbb{Z}/k\mathbb{Z}$ defined in in Example **??** is just the quotient set $\mathbb{Z}/k\mathbb{Z}$. Check that $\mathbb{Z}/k\mathbb{Z} = \{[0], [1], \dots, [k-1]\}$. Thus $[\mathbb{Z} : k\mathbb{Z}] = k$.

THEOREM 2.20 (Lagrange). *Let G be a finite group. Then $|H|$ divides $|G|$. More precisely $|G| = |H| \cdot [G : H]$.*

PROOF. By the previous lemma and by $H = 1H$, we see that G is the disjoint union of $[G : H]$ different cosets aH with $|H|$ elements each, thus $|G| = |H|[G : H]$. □

Although the proof is not difficult, this theorem is very important and very useful, it will be used many times in the future. For the moment we will illustrate Lagrange's theorem by giving a number of corollaries. First we introduce a definition.

DEFINITION 2.21. Let G be a group and $a \in G$. The *order* of a, denoted $ord(a)$ is the smallest positive integer n, such that $a^n = 1$, if such an n exists, otherwise $ord(a)$ is defined to be ∞.

For instance in any group $ord(1) = 1$, and for any $n \in (\mathbb{Z}, +) \setminus \{0\}$, we have $ord(n) = \infty$.

COROLLARY 2.22. *If G is a finite group and $a \in G$, then $ord(a)$ divides $|G|$.*

PROOF. Consider the cyclic subgroup $\langle a \rangle$ of G generated by a. By Lagrange's Theorem the Corollary follows immediately from the following

Claim $|\langle a \rangle| = ord(a)$**.**

Proof of the Claim: As G is finite, also $\langle a \rangle$ is finite. We claim that $\langle a \rangle = \{1, a, a^2, \dots, a^{ord(a)}\}$, and these elements are distinct. By definition $\langle a \rangle = \{a^n \mid n \in \mathbb{Z}\}$. By division with rest in $\mathbb{Z}$, we can write $n = ord(a) \cdot d + r$ with $r, d \in \mathbb{Z}$ and $0 \leq r < ord(a)$.

Then $a^n = a^{ord(a) \cdot d + r} = (a^{ord(a)})^d a^r = 1^d a^r = a^r$. Finally we have to see that the a^r with $0 \leq r < ord(a)$ are distinct. Assume $0 \leq r_2 < r_1 < ord(a)$ with $a^{r_1} = a^{r_2}$, then $a^{r_1 - r_2} = 1$ and $0 < r_1 - r_2 < ord(a)$ contradicting the definition of $ord(a)$. □

COROLLARY 2.23. *If G is a finite group and $a \in G$, then $a^{|G|} = 1$.*

PROOF. By the previous Corollary $ord(a)$ divides $|G|$, i.e we can write $|G| = ord(a) \cdot d$. Thus $a^{|G|} = (a^{ord(a)})^d = 1^d = 1$. □

COROLLARY 2.24. *Let G be a finite group of order a prime number p, then G is a cyclic group.*

PROOF. p has 1 and p as the only nonnegative divisors. Note that G has no subgroups H other then $\{1\}$ and G, because otherwise $|H|$ would be a divisor of $|G|$ different from 1 and p. Let $a \neq 1 \in G$. Then $\langle a \rangle$ is a subgroup of G different from $\{1\}$. Therefore $\langle a \rangle = G$. □

3. Normal subgroups and quotient groups

If H is a subgroup of a group G, we can form the set of cosets

$$G/H = \big\{ aH \;\big|\; a \in G \big\}.$$

G is a group, does G/H inherit the group structure from G? There is an obvious definition for the product in G/H:

$$(aH)(bH) := abH,$$

but does it make sense? For this definition to make sense we have to remember that there can be many different $a' \in G$, $b' \in G$ with $a'H = aH$, $b'H = bH$, but above we want to define the product of $(aH)(bH)$ in terms of the actual elements ab. This makes sense precisely when $a'b'H = abH$ whenever $a'H = aH$, $b'H = bH$. In other words we need the product to be independent of the representatives. Or in other words we need it to be well-defined.

We now want to find a necessary condition for the product to be well defined (later we will see it is also sufficient). For this we chose $a = 1$, $a' = h$ an arbitrary element of H, b an arbitrary element of G and $b' = b$. Thus we want to see under what circumstances we will always have $bH = hbH$. This is equivalent to $hb = bh'$ for some element $h' \in H$. Multiplying by b^{-1} this is equivalent to $b^{-1}hb = h'$ for some element $h' \in H$, i.e. $b^{-1}hb \in H$. As h was an arbitrary element in H, b an arbitrary element in G, we find that a necessary condition for the multiplication above to be well-defined is $b^{-1}hb \in H$ for all $h \in H$ and all $b \in G$.

This is an extra condition on H. Subgroups H fulfilling this condition are called normal subgroups.

DEFINITION 3.1. Let G be a group, and let $H \subset G$ be a subgroup. H is called a *normal subgroup* of G if for all $g \in G$, and all $h \in H$ we have $ghg^{-1} \in H$.

EXAMPLE 3.2.
(1) If G us a commutative group, then every subgroup is a normal subgroup ($ghg^{-1} = h$).
(2) Let G be a group. Then $\{1\}$ and G are normal subgroups of G.
(3) In S_3, consider

$$H := \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = e, \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \right\}.$$

Note that

$$\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}=\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}=e,$$

thus H is a subgroup of S_3, and $\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}=\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}^{-1}$. On the other hand one computes

$$\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}=\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix},$$

therefore H is not a normal subgroup of S_3.

Now we want to see that if N is a normal subgroup of G, we get indeed that with the above definition G/N is a group.

THEOREM 3.3. *Let G be a group and let $H \subset G$ be a normal subgroup. Then G/H is a group with the product $(aH)(bH) = abH$.*

PROOF. We have first show that the product is well-defined. Let $a, b, a', b' \in G$ and assume $aH = a'H$, $bH = b'H$. We have to see $abH = a'b'H$. Equivalently assume $a' \equiv a \mod H$, $b' \equiv b \mod H$. We have to see that $ab \equiv a'b' \mod H$. By assumption we have $a' = ah_1$, $b' = bh_2$ with $h_1, h_2 \in H$. Then $a'b' = ah_1bh_2 = ab(b^{-1}h_1b)h_2$. As H is a normal subgroup, we have $(b^{-1}h_1b) \in H$ and therefore also $(b^{-1}h_1b)h_2 \in H$. Therefore $a'b' \equiv ab \mod H$.

The group axioms for this product follow directly from the group axioms for G: Associativity:

$$((aH)(bH))cH = ((ab)H)(cH) = (ab)cH = a(bc)H = aH(bcH) = aH((bH)(cH)),$$

Neutral element: $(1H)(aH) = 1aH = aH$. Inverse: $(aH)(a^{-1}H) = aa^{-1}H = 1H$. □

4. Group homomorphisms

An important concept in modern algebra (and in modern mathematics) is the notion of homomorphism. If one has sets G_1, G_2 with a structure (e.g. they are groups), a homomorphism will a map $f : G_1 \to G_2$ which is compatible with the structure. For groups this means that the map must be compatible with the multiplication.

DEFINITION 4.1. Let G, H be groups and let $f : G \to H$ be a map. f is called a homomorphism if

$$f(ab) = f(a)f(b)$$

for all $a, b \in G$.

Examples of group homomorphisms

(1) Let G be a group, then $G \to \{1\}, g \mapsto 1$ is a group homomorphism to the trivial group.

(2) Let $k \in \mathbb{Z}$ be an integer. Then $\cdot k : \mathbb{Z} \to \mathbb{Z},\ n \mapsto kn$ is a group homomorphism: $k(n+m) = kn + km$.
(3) The exponential map $exp : \mathbb{R} \to \mathbb{R} \setminus \{0\};\ a \mapsto e^a$ is a homomorphism from $(\mathbb{R}, +)$ to $(\mathbb{R}^*, \cdot)$: $e^{a+b} = e^a e^b$.
(4) The map $\pi : \mathbb{Z} \to \mathbb{Z}_k; n \mapsto \underline{n}$, is a group homomorphism (exercise).

REMARK 4.2. Let $f : G \to H$ be a group homomorphism.
Let e be the neutral element of G and e' the neutral element of H. Then $f(e) = e'$. $f(a^{-1}) = f(a)^{-1}$ for all $a \in G$.

PROOF. (1) $e'f(e) = f(e) = f(ee) = f(e)f(e)$ by cancelation $e' = f(e)$. (2) $e' = f(e) = f(a^{-1}a) = f(a^{-1})f(a)$, therefore $f(a)^{-1} = f(a^{-1})$. □

The composition of group homomorphisms is a group homomorphism

REMARK 4.3. Let $f : G \to H$, $g : H \to L$ be group homomorphisms. Then $g \circ f : G \to L$ is a group homomorphism.

PROOF. We have for all $a, b \in G$: $g \circ f(ab) = g(f(a)f(b)) = g(f(a))g(f(b))$. □

DEFINITION 4.4. Let $f : G \to H$ be a group homomorphism. The set

$$ker(f) := \{a \in G \mid f(a) = 1\}$$

is called the *kernel* of f. The set

$$Im(f) := \{b \in H \mid \text{ there is an } a \in G \text{ with } f(a) = b\}$$

is called the *image* of f.

PROPOSITION 4.5. *Let $f : G \to H$ be a homomorphism of groups.*
(1) *The image $Im(f)$ is a subgroup of H.*
(2) *The kernel $ker(f)$ is a normal subgroup of G.*

The second statement will be particularly important.

PROOF. (1) is quite obvious: $Im(f) \subset H$ and for $h_1 = f(a), h_2 = f(b) \in Im(f)$ we also have $h_1h_2 = f(a)f(b) = f(ab) \in Im(f)$, furthermore for $h = f(a) \in f(a)$ also $h^{-1} = f(a)^{-1} = f(a^{-1})$ in $Im(f)$.
(2) $ker(f)$ is a subgroup: If $a, b \in ker(f)$, then $f(a) = f(b) = 1$. Thus $f(ab) = f(a)f(b) = 1 \cdot 1 = 1$, thus $ab \in ker(f)$, and $f(a^{-1}) = f(a)^{-1} = 1^{-1} = 1$, so $a^{-1} \in ker(f)$.
Finally we show $ker(f)$ is a normal subgroup. Let $h \in ker(f)$, $a \in G$. Then $f(a^{-1}ha) = f(a^{-1})f(h)f(a) = f(a)^{-1}f(a) = 1$. thus $a^{-1}ha \in ker(f)$. □

DEFINITION 4.6. A homomorphism $f : G \to H$ is called
isomorphism, if f is bijective.
We will call two groups G, H isomorphic, if there is an isomorphism from G to H. The homomorphisms $f : G \to G$ are also called *endomorphisms* of G, and the isomorphims $f : G \to G$ are also called *automorphisms*.

EXAMPLE 4.7. (1) For $\cdot k : \mathbb{Z} \to \mathbb{Z}$, $n \mapsto kn$ we have $ker(\cdot k) = \{0\}$, $Im(\cdot k) = k\mathbb{Z} := \{kn \mid n \in \mathbb{Z}\}$. Thus $\cdot k$ is injective, but not an isomorphism.
(2) For $exp : (\mathbb{R}, +) \to (\mathbb{R}_{>0}, \cdot)$ is an isomorphism.

REMARK 4.8. (1) Let $f : G \to H$ be a group isomorphism, then also $f^{-1} : H \to G$ is an isomorphism.
(2) Clearly if $\varphi : G \to H$, $\psi : H \to L$ are isomorphisms, then $\psi \circ \varphi : G \to L$ is an isomorphism.

PROOF. (1) Let $a', b' \in H$. Then there are unique $a, b \in G$ with $f(a) = a'$, $f(b) = b'$. Thus we have

$$f^{-1}(a'b') = f^{-1}(f(a)f(b)) = f^{-1}(f(ab)) = ab = f^{-1}(a')f^{-1}(b').$$

Therefore f^{-1} is a homomorphism, and as it is also bijective it is an isomorphism. □

DEFINITION 4.9. We say that groups G and H are isomorphic (denoted $G \simeq H$ if there is an isomorphism $\varphi : G \to H$.

If is clear that being isomorphic is an equivalence equivalence relation on groups, i.e.

(1) $G \simeq G$,
(2) $G \simeq H$ implies $H \simeq G$,
(3) $G \simeq H$ and $H \simeq L$ imply that $G \simeq L$.

That G and H are isomorphic means that they have precisely the same properties as groups, and for most purposes we can identify them.

If $f : G \to H$ is a morphism of groups, then $Im(f)$ is subgroup of H, thus we can replace f by the *surjective* group homomorphism $f : G \to Im(f)$.

We can check whether a surjective morphism is an isomorphism by just looking at the kernel.

LEMMA 4.10. (1) *Let $f : G \to H$ be a group homomorphism. Then f is injective if and only if $ker(f) = \{1\}$.*
(2) *In particular a surjective group homomorphism $f : G \to H$ is an isomorphims if and only if $ker(f) = \{1\}$.*

PROOF. Let e be the neutral element of H.

"$\Longrightarrow$" We know $f(1) = e$, therefore $ker(f) \supset \{1\}$. As f is injective $f(a) \neq e$ for all $a \neq 1$. Thus $\ker(f) = \{1\}$.

"$\Longleftarrow$" Assume $ker(f) = \{1\}$. Let $a, b \in G$, assume $f(a) = f(b)$. Then $f(a^{-1}b) = f(a^{-1})f(b) = f(a)^{-1}f(b) = e$, thus $a^{-1}b \in ker(f) = \{1\}$, Therefore $a = b$. Thus f is injective. □

EXERCISE 4.11. Let $f : M \to N$ be a bijection of sets. The map

$$f^* : S(N) \to S(M), \sigma \mapsto f \circ \sigma \circ f^{-1}$$

is an isomorphism of groups. In particular if M is a finite set with n elements then $S(M) \simeq S_n$.

The following lemma gives us many surjective group homomorphisms. We will see in a moment that in a suitable way it gives us the most general example of a surjective group homomorphism.

LEMMA 4.12. *Let G be a group, and let N be a normal subgroup of G. The natural map $\pi : G \to G/N; a \to aN$ is a surjective group homomorphism, with kernel N.*

PROOF. By definition for $a, b \in G$ we have $\pi(a)\pi(b) = (aN)(bN) = abN = \pi(ab)$, so π is a group homomorphism. If $a \in ker(\pi)$, then $aN = N$, in particular $a = a1 \in N$. Conversely if $a \in N$, then $\pi(a) = aN = N = 1N$. Thus $ker(\pi) = N$. □

If G is a group and N is a normal subgroup, we thus get a surjective homomorphism $G \to G/H$ with kernel H. We can hope to understand G in terms of H and G/H which are both smaller, (and therefore maybe simpler). Iterating this process we end up with groups G which have no normal subgroups except $\{1\}$ and G

DEFINITION 4.13. A group G is called *simple* if its only normal subgroups are $\{1\}$ and G.

Simple groups can be viewed as the basic building blocks of all groups. One of the most difficult theorems in mathematics is the classification of all finite simple groups. That is a list of all finite simple groups up to isomorphism.

EXAMPLE 4.14. Clearly $\mathbb{Z}/2\mathbb{Z}$ is simple.

THEOREM 4.15. *(Homomorphism Theorem) Let $\varphi : G \to H$ be a surjective morphism, with kernel K. Then there is an isomorphism $\overline{\varphi} : G/K \to H$ with*

$$\overline{\varphi} \circ \pi = \varphi.$$

In particular H is isomorphic to $G/ker(\varphi)$.

PROOF. The condition $\varphi = \overline{\varphi} \circ \pi$, says that for every $a \in G$ we have $\overline{\varphi}(aK) = \varphi(a)$, we want to use this as definition for $\overline{\varphi}$. We have to see that this is well defined. That is we have to see that if $aK = bK$, then $\varphi(a) = \varphi(b)$. But $aK = bK$ means $a \equiv b \mod K$, i.e. $a = bk$ for some $k \in K$. Then $\varphi(a) = \varphi(bk) = \varphi(b)\varphi(k) = \varphi(b)1 = \varphi(b)$. Thus φ is well-defined. Clearly $\overline{\varphi}$ is a homomorphism because $\overline{\varphi}((aK)(bK)) = \overline{\varphi}(abK) = \varphi(ab) = \varphi(a)\varphi(b) = \overline{\varphi}(aK)\overline{\varphi}(bK)$.

To show that $\overline{\varphi}$ is an isomorphism, we have to show that $ker(\overline{\varphi}) = 1K$. By definition $aK \in ker(\overline{\varphi})$ if and only if $\varphi(a) = 1$, i.e. if and only if $a \in K$. And we know that $aK = 1K$ if and only if $a \in K$. □

EXAMPLE 4.16. Let k be an integer.

(1) $k\mathbb{Z}$ is a subgroup of $(\mathbb{Z}, +)$, and, as $\mathbb{Z}$ is commutative, it is a normal subgroup. Thus we can form the quotient group $\mathbb{Z}/k\mathbb{Z}$. Note that this definition coincides precisely with that of Example **??**. We have seen that the map $n \mapsto \underline{n}$ defines a surjective group homomorphism $\mathbb{Z} \to \mathbb{Z}_k$ with kernel $k\mathbb{Z}$, thus the map $[n] \mapsto \underline{n}$ defines a group isomorphism $\mathbb{Z}/k\mathbb{Z} \to \mathbb{Z}_k$.
(2) Let G be a cyclic group of order k. Then G is isomorphic to $\mathbb{Z}/k\mathbb{Z}$. Let a be generator of G. Then it is easy to see that $[n] \to a^n$ defines an isomorphism $\mathbb{Z}/k\mathbb{Z} \to G$.

Later we will use the following elementary lemma.

LEMMA 4.17. *Let G be a finite abelian group of finite order nm, and let H, L be subgroups with $|H| = n$. $|L| = m$ and $L \cap M = \{1\}$. Then G is isomorphic to $H \times L$.*

PROOF. We define a homomorphism $\psi : H \times L \to G; (h, l) \mapsto hl$. This is a homomorphism because

$$\psi((h, l)(h', l')) = \psi(hh', ll') = hh'll' = hlh'l' = \psi(h, l)\psi(h', l').$$

We claim the kernel is $\{(1, 1)\}$. Let $(h, l) \in ker(\psi)$, then $hl = 1$, therefore $l = h^{-1}$. But as H is a subgroup of G, we have $h^{-1} \in H$, thus $l \in H \cap L$, therefore $l = 1$, and therefore also $h = 1$. Thus ψ is an injective group homomorphism. It is bijective because G and $H \times L$ have the same number of elements. □

Automorphisms

An automorphism φ of a group G is an isomorphism $\varphi : G \to G$. We denote by $Aut(G)$ the set of automorphisms of G. We now want to show that $Aut(G)$ is a group with composition of maps as multiplication. An automorphism is in particular a bijective map $\varphi : G \to G$, therefore $Aut(G)$ is a subset of the symmetric group $S(G)$. We know that $S(G)$ is a group with multiplication the composition of maps $\varphi \cdot \psi := \varphi \circ \psi$, the neutral element being the identity map $G \to G$.

LEMMA 4.18. *$Aut(G)$ is a subgroup of $S(G)$.*

PROOF. Let $f, g \in Aut(G)$, we have to show that $f \circ g \in Aut(G)$ and that $f^{-1} \in Aut(G)$. But we know that the composition of isomorphisms is an isomorphism, and that the inverse of an isomorphism is an isomorphism, so this is clear. □

There is a special class of automorphism of every group, the inner automorphisms.

DEFINITION 4.19. Let G be a group. Then for every $a \in G$ the map $\tau_a : G \to G, b \mapsto aba^{-1}$ is an automorphism. (It is clear that this is a homomorphism, because for $g, a, b \in G$, we have $\tau_g(ab) = gabg^{-1} = (gag^{-1})(gbg^{-1}) = \tau_g(a)\tau_g(b)$. Furthermore they are injective: if $a \in ker(\tau_g)$, then $gag^{-1} = 1$, thus $ga = g$ and thus $a = 1$. And they are surjective: for $b \in G$ we find $\tau_g(g^{-1}bg) = b$.)

An automorphism of G is called an *inner automorphism*, if it is of the form τ_a for some $a \in G$. We denote by $inn(G)$ the group of inner automorphisms of G.

If G is an abelian group then all inner automorphisms are equal to the identity $id : G \to G$.

PROPOSITION 4.20. *Let G be a group. Then $inn(G)$ is a normal subgroup of $Aut(G)$.*

PROOF. First we show $inn(G)$ is a subgroup. For this we have to see that for $a, b \in G$ we have $\tau_a \circ \tau_b$ is an inner automorphism. But by definition for any $x \in G$ we have $\tau_a \circ \tau_b(x) = \tau_a(bxb^{-1}) = (ab)x(ab)^{-1} = \tau_{ab}(x)$. Thus $\tau_a\tau_b = \tau_{ab}$. Similarly $\tau_a\tau_{a^{-1}} = id_G$, thus $(\tau_a)^{-1} = \tau_{a^{-1}}$.

Furthermore let $\varphi : G \to G$ be an isomorphism, then for any $a, x \in G$ we have $\varphi \circ \tau_a \circ \varphi^{-1}(x) = \varphi(a\varphi^{-1}(x)a^{-1}) = \varphi(a)x\varphi(a^{-1}) = \varphi(a)x\varphi(a)^{-1} = \tau_{\varphi(a)}(x)$, i.e. we get $\varphi \circ \tau_a \circ \varphi^{-1} = \tau_{\varphi(a)} \in inn(G)$. □

It is actually not difficult to describe $inn(G)$ as a group, it is a quotient group of G, by its center, the subgroup of all elements of G which commute with all other elements.

DEFINITION 4.21. Let G be a group. The *center* of G, denoted $Z(G)$ is the set

$$Z(G) := \big\{a \in G \;\big|\; ah = ha, \text{ for all } h \in G\big\}.$$

It is easy to see that $Z(G)$ is a normal subgroup of G: If for all $h \in G$ we have $ah = ha$ and $bh = hb$, then we have $abh = ahb = hab$, i.e. if $a, b \in Z(G)$, then $ab \in Z(G)$. Similarly, if for all $h \in G$ $ah = ha$, then multiplying on both sides by a^{-1} we get $ha^{-1} = a^{-1}h$, i.e. $a^{-1} \in Z(G)$. Thus $Z(G)$ is a subgroup of G. Finally if $a \in Z(G)$, and $h \in G$, then $hah^{-1} = ahh^{-1} = a \in Z(G)$, thus $Z(G)$ is a normal subgroup.

PROPOSITION 4.22. *$inn(G) \simeq G/Z(G)$.*

PROOF. We have seen above that $\tau_a\tau_b = \tau_{ab}$. Therefore the map $\tau : G \to inn(G)$, $a \mapsto \tau_a$ is a group homomorphism. By definition τ is surjective. We determine its kernel: We have $a \in ker(\tau)$ if and only if $\tau_a = id_G$. This is equivalent to $\tau_a(x) = x$ for all $x \in G$, i.e. $axa^{-1} = x$. Multiplying by a on the right this is equivalent to $ax = xa$ for all $x \in G$. In other words $ker(\tau) = Z(G)$. So the result follows by the homomorphism theorem. □

5. Group operations

As we mentioned earlier, groups are related to the concept of symmetry. Intuitively a symmetry of an object is a way to move it around with out changing its form, e.g. a square can be reflected at the lines though the middle of the sides and at the diagonals, and it can also be turned around by 90 degrees.

Usually this symmetry is given by the operation of a group, which we now want to explain. Given a set X and a group G, an action of G on X is a way to to "multiply"

elements of X with elements of G, i.e. for every $g \in G$ and every $x \in X$ we have $g \cdot x \in X$, such that

(1) The neutral element $1 \in G$ acts as identity on X, i.e. $1 \cdot x = x$ for all $x \in X$,
(2) the operation is compatible with the group structure on X, i.e. there is an "associative law" $(ab) \cdot x = a \cdot (b \cdot x)$ for all $a, b \in G$, $x \in X$.

We write this again as a formal definition.

DEFINITION 5.1. Let G be a group, and let X be a set. An *operation* (also called *action*) of G on X is a map

$$\cdot : G \times X \to X;\ (g, x) \mapsto g \cdot x,$$

so that the following axioms are satisfied.

(1) $1 \cdot x = x$ for all $x \in X$,
(2) $(ab) \cdot x = a \cdot (b \cdot x))$ for all $a, b \in G$, $x \in X$. (On the left hand side the leftmost product is the product in G, $\cdot$ always denotes the operation of G on X).

EXAMPLE 5.2.
(1) Let X be a nonempty set, G a group. The *trivial operation* of G on X is defined by $g \cdot x = x$ for all $g \in G$, $x \in X$.
(2) Let S_n the symmetric group in n letters. Then S_n acts on $\{1, \dots, n\}$ by $\sigma \cdot k := \sigma(k)$; ($\sigma$ is a map of $\{1, \dots, n\}$ onto itself and we apply it to $k \in \{1, \dots, n\}$). Recall that the neutral element e of S_n is the identity map $id : \{1, \dots, n\} \to \{1, \dots, n\}$, Obviously $e \cdot k = k$. By definition of the composition of maps we have $(\sigma\tau \cdot k = \sigma \circ \tau(k) = \sigma(\tau(k)) = \sigma \cdot (\tau \cdot k)$, so this is a group operation.
(3) Let G be a group. Then multiplication in G defines an operation $l : G \times G \to G$, $(g, h) \mapsto gh$, the left translation. This is clearly an operation of G on itself.
(4) Let G be a group, the map $G \times G \to G$, $(g, h) = ghg^{-1}$ is an operation of G on itself. It is called *conjugation*.

We find that every element $g \in G$ acts as a bijection of X onto itself:

DEFINITION 5.3. Let G be a group operating on a nonempty set X. Let $g \in G$. The *multiplication by* g is the map

$$m_g : X \to X; x \mapsto g \cdot x.$$

Note that by definition $m_1 = id_X$. And we have $m_g \circ m_{g^{-1}}(x) = (g \cdot g^{-1}) \cdot x = x$ and similarly $m_{g^{-1}} \circ m_g = id_X$. Thus for all $g \in G$, m_g is a bijection $m_g \in S(X)$.

LEMMA 5.4. *Let $\cdot : G \times X \to X$ be a group operation. Then the map $m : G \to S(X)$, $g \mapsto m_g$ is a group homomorphism.*

PROOF. Let $g, h \in G$, $x \in X$, then by definition we get $m_{gh}(x) = (gh) \cdot x = g \cdot (h \cdot x) = (m_g \circ m_h)(x)$. As the multiplication in $S(X)$ is the composition, this says that m is a group homomorphism. □

We use this to prove that every finite group is isomorphic to a subgroup of some symmetric group S_n.

THEOREM 5.5. *(Cayley's theorem) Every finite group G is isomorphic to a subgroup of some symmetric group S_n. More precisely if G has m elements, then G is isomorphic to a subgroup of S_m.*

PROOF. Let $m = |G|$. Recall that the symmetric group $S(G)$ is isomorphic to S_m. We apply the previous lemma to the multiplication operation $\cdot : G \times G \to G, (g, h) \mapsto gh$. Then $m : G \to S(G); g \mapsto m_g$ is a group homomorphism. We only need to show that m is injective. Then G is isomorphic to $im(m)$, a subgroup of $S(G) \simeq S_m$. Let $g \in ker(m)$, then $m_g = id_G$, i.g. $gh = h$ for all $h \in G$. By the cancelation property, we get $g = 1$. Thus $ker(m) = 1$ and m is injective. □

This allows us to describe every finite group up to isomorphism as a group of permutations. This is indeed used in some computer algebra programs to make computations in groups. There is however a price: when G has m elements, G is described as a subgroup of S_m which has $m!$ elements, so this description is not always the most efficient one.

EXAMPLE 5.6. (the dihedral group). The dihedral group D_n is the group of symmetries of a regular n-gon by rotations and reflections. We can view it as a subgroup of S_n via its action on the n vertices of the n-gon. We number the vertices cyclically with $1, \dots, n$.

Then the rotation by the angle $2\pi/n$ will permute the vertices via

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ 2 & 3 & 4 & \dots & 1 \end{pmatrix}$$

and in case n is odd

$$\tau = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ 1 & n & n-1 & \dots & 2 \end{pmatrix}$$

is the reflection on the axis through the vertex one and the opposite side, or in case n is even

$$\tau = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ n & n-1 & n-2 & \dots & 1 \end{pmatrix}$$

is the reflection on the axis orthogonal to the side through 1 and n. In any case the dihedral group D_n is defined as the subgroup of S_n generated by σ and τ. It is easy to see that $ord(\sigma) = n$, $ord(\tau) = 2$, and that $\tau\sigma\tau = \sigma^{-1}$. From this is is easy to check that

$$D_n = \left\{ \sigma^i \tau^j \;\middle|\; 0 \le i \le n-1,\ 0 \le j \le 1 \right\}.$$

And in particular $|D_n| = 2n$.

DEFINITION 5.7. Let G be a group with an operation on a nonempty set X. The operation is called *transitive* if, for every x, y in X there is a $g \in G$ with $gx = y$. It is called *simply transitive* if, for every x, y in X there is a unique $g \in G$ with $gx = y$.

EXERCISE 5.8. Let G be a group acting simply transitively on a set X. Then for any $x \in X$ the map $g \mapsto g \cdot x$ is a bijection $G \to X$.

We now introduce some fundamental notions connected to a group operation. The orbit of an element and the stabilizer of an element.

The orbit of $x \in X$ consists of all elements of X that can be reached by operating with elements of G on x.

DEFINITION 5.9. Let a group G operate on a nonempty set X. Let $x \in X$. The *orbit* of x is

$$G(x) := \big\{ g \cdot x \;\big|\; g \in G \big\} \subset X.$$

A subset $Y \subset X$ is called an *orbit* of G on X, if it is of the form $Y = G(x)$ for some $x \in X$.

We see that the action of G on X is transitive if and only if $G(x) = X$ for any $x \in X$. On the other hand for any action, restricting the action of G to $G(x)$, we get a transitive action of G on $G(x)$.

The group action of G on X defines an equivalence relation of X whose equivalence classes are the orbits of G on X.

DEFINITION 5.10. We define an equivalence relation $\sim$ on X, by $x \sim y$, if and only if there is a $g \in G$ with $g \cdot x = y$. It is easy to see that $\sim$ is an equivalence relation, and that the equivalence classes are precisely the orbits.

The set of equivalence classes (which is equal to the set of orbits of the action of G on X) is called the *orbit space* and denoted X/G.

The stabilizer of an element $x \in X$ is the subgroup of all $g \in G$ which fix x, i.e. $g \cdot x = x$.

DEFINITION 5.11. Let a group G act on a nonempty set X. The *stabilizer* of x is

$$G_x := \big\{ g \in G \;\big|\; g \cdot x = x \big\}.$$

It is obvious that G_x is a subgroup of X.

LEMMA 5.12. *Let a group G act on a nonempty set X. Let $x \in X$. Then there is a bijection $G/G_x \to G(x)$. In particular if $G(x)$ is finite $|G(x)| = [G : G_x]$.*

PROOF. The map $G/G_x \to G(x), gG_x \to g \cdot x$ is well-defined: If $g' = gh$ with $h \in G_x$, then $g' \cdot x = g \cdot (h \cdot x) = g \cdot x$. It is also injective: if $g \cdot x = g' \cdot x$, then $(g^{-1}g') \cdot (x) = g^{-1} \cdot (g' \cdot x) = g^{-1} \cdot g \cdot x = x$. Thus $h := g^{-1}g' \in G_x$ and $g' = gh$, thus $gG_x = g'G_x$. The map is surjective by definition, thus it is a bijection. □

DEFINITION 5.13. Let $\sim$ be an equivalence relation on a set X. A subset $R \subset X$ is called a system of representatives for $\sim$, if it contains precisely one element from every equivalence class.

THEOREM 5.14. *(Orbit stabilizer theorem) Let a group G act on a nonempty finite set X. If R is a system of representatives for the orbits, then*

$$|X| = \sum_{x \in R} |G(x)| = \sum_{x \in R} [G : G_x].$$

PROOF. We know that X is the disjoint union of the orbits $G(x)$, so the first equality is clear. The second follows by the previous Lemma. □

.

Action by conjugation In this section we will study the action of G on itself by conjugation, $(g, h) \mapsto ghg^{-1}$.

DEFINITION 5.15. Let G be a group. Let $g, h \in G$. We say g is conjugated to h if there is an element $a \in G$ with $aga^{-1} = h$. It is straightforward to see that this is an equivalence relation.

The equivalence classes $[g]$ of G with respect to conjugation are called the conjugacy classes of G.

DEFINITION 5.16. Let G be a group and $x \in G$. The *centralizer* of x in G is the set

$$Z(x) := \{g \in G \mid gx = xg\}.$$

REMARK 5.17.
(1) It is immediate from the definition that $Z(x)$ is a subgroup of G.
(2) Also by definition the Center $Z(G)$ of G is a subgroup of $Z(x)$ for any $x \in G$.
(3) Let $x \in G$. By definition $x \in Z(G)$ if and only if $Z(x) = G$.

The conjugacy classes contain important information about the group, and it is therefore important to study them. We will do this via the action of G on itself by conjugation.

DEFINITION 5.18. The conjugation action of G on itself is defined by $g \cdot h := ghg^{-1}$. By definition the orbits of the conjugation action are precisely the conjugacy classes, i.e. $G(g) = [g]$.

An important result is the

THEOREM 5.19. *(Class equation). Let G be a finite group, let $Z(G)$ be the center of G and let R be a set such that every element of $G \setminus Z(G)$ is conjugated to precisely one element of R. Then*

$$|G| = |Z(G)| + \sum_{x \in R} [G : Z(x)].$$

PROOF. Being conjugated is an equivalence relation. Thus if R' is a set of representatives for this equivalence relation we get $|G| = \sum_{x \in R'} |[x]| = \sum_{x \in R'} |G(x)|$ for the conjugation action. By definition the stabiliser of $x \in G$ under the conjugation action is $Z(x)$. Thus we get by the orbit stabilizer theorem

$$|G| = \sum_{x \in R'} |[x]| = \sum_{x \in R'} [G : Z(x)].$$

Note that if $y \in Z(G)$, then $[y] = \{y\}$, thus $Z(G)$ is a subset of R. Put $R = R' \setminus Z(G)$, and note $[G : Z(x)] = 1$ for $x \in Z(G)$. Then

$$|G| = \sum_{x \in Z(G)} 1 + \sum_{x \in R} [G : Z(x)],$$

and the claim follows. □

We first introduce a concept which will be important in the future.

DEFINITION 5.20. Let p be a prime number. A p-group is a group of order a power p^n for $n \in \mathbb{Z}_{>0}$.

Let p be a prime number. We know that if G is a finite group with $|G| = p$, then G is cyclic, in particular it is abelian. Now we want to show that also groups of order p^2 are abelian.

PROPOSITION 5.21. *The center of a p-group has order at least p.*

PROOF. Let G be a p-group of order p^n. Then $Z(G)$ is a subgroup of G, and thus $|Z(G)|$ divides p^n. Therefore it is enough to show that $Z(G)$ contains an element different from 1. Assume $Z(G) = \{1\}$. Then the class equation reads

$$p^n = |G| = 1 + \sum_{x \in R} [G : Z(x)],$$

but as $Z(x)$ is a subgroup of G different from G, we get $[G : Z(x)]$ is a positive power of p for all $x \in R$. Thus the left hand side of the equation is divisible by p and the right hand side is not. This is a contradiction. □

PROPOSITION 5.22. *Let p be a prime number. Then every group of order p^2 is abelian.*

PROOF. Let G be a group of order p^2. We want to show that for every $x \in G$ the centralizer $Z(x)$ is the whole group G. Then it follows that G is abelian. If $x \in Z(G)$, then by definition $Z(x) = G$. If $x \notin Z(G)$, then $Z(x)$ contains $Z(G)$ and x. Thus it contains at least $p + 1$ elements. As $Z(x)$ is a subgroup of G the number $|Z(x)|$ must be a divisor of p^2, and thus $Z(x) = G$. □

6. The symmetric group

We now want to give an explicit description of the symmetric group S_n. We have seen that every finite group is a subgroup of some S_n.

The elements of S_n are the bijections $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$. They form a group under composition $\sigma\tau = \sigma \circ \tau$. We have earlier introduced the following notation for elements in S_n: we write the permutation σ as a matrix

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}.$$

We now want to introduce another description of permutations: in terms of a cycle decomposition.

DEFINITION 6.1. Let $a_1, \dots, a_r$ be r distinct elements of $\{1, \dots, n\}$. The cycle $(a_1, \dots, a_r)$ is the permutation σ given by $\sigma(a_i) = a_{i+1}$ for all $i = 1, \dots, r-1$, $\sigma(a_r) = a_1$ and $\sigma(k) = k$ if $k \notin \{a_1, \dots, a_r\}$. We call r the *length* of the cycle $(a_1, \dots, a_r)$.

Note that the notation is not unique: different r-tuples define the same cycle: For any $l \in \{1, \dots, r\}$ we have $(a_l, a_{l+1}, \dots, a_r, a_1, \dots, a_{l-1})$ and $(a_1, \dots, a_r)$ are the same cycles. Cycles of length n are called n-cycles. Cycles of length 2 are called *transpositions*. Note that any (1)-cycle (k) is just the identity element of S_n.

Two cycles $(a_1, \dots, a_r)$, $(b_1, \dots, b_s)$ are called *disjoint* if the sets $\{a_1, \dots, a_r\}$ and $\{b_1, \dots, b_s\}$ are disjoint.

THEOREM 6.2. *Let $n \geq 2$.*

1. *Every element $\sigma \in S_n$ is the product $\sigma = \sigma_1 \dots \sigma_s$ or pairwise disjoint cycles for some $s \geq 0$.*
2. *Any two disjoint cycles σ, τ commute: $\sigma\tau = \tau\sigma$.*
3. *Every $\sigma \in S_n$ is a finite product of transpositions.*

PROOF. (1) Let $H = \langle \sigma \rangle$ be cyclic subgroup of S_n generated by σ. The map

$$\cdot : H \times \{1, \dots, n\} \to \{1, \dots, n\}, (\tau, i) \mapsto \tau \cdot i := \tau(i)$$

is an operation of H on $\{1, \dots n\}$. Choose $b_1, \dots, b_r \in \{1, \dots, n\}$, so that $H(b_1), \dots, H(b_r)$ are the different orbits of H. For each $i = 1, \dots, r$ let m_i be $min(k \in \mathbb{Z}_{>0} | \sigma^k(b_i) = b_i)$. Then we see that $b_i, \sigma(b_i), \dots, \sigma^{m_i - 1}(b_i)$ are pairwise distinct, and that $H(b_i) = \{b_i, \sigma(b_i), \dots, \sigma^{m_i-1}(b_i)\}$. Thus the $\sigma_i := (b_i, \sigma(b_i), \dots, \sigma^{m_i-1}(b_i))$ are for $i = 1, \dots, r$ disjoint cycles. Furthermore we have $\sigma_1 \circ \dots \circ \sigma_r = \sigma$, because each $x \in \{1, \dots, n\}$ lies precisely in one of the $H(a_i)$: Thus for x there is precisely one $i \in \{1, \dots, r\}$ and one $k \in \{1, \dots, m_i\}$ with $x = \sigma^k(a_i)$. Then $\sigma(x) = \sigma_i(x)$ and $\sigma_j(x) = x$ for all $j \neq i$. Thus $\sigma_1 \dots \sigma_r(x) = \sigma_i(x) = \sigma(x)$.

(2) is obvious. (3) Note that $(a_1, \dots, a_r) = (a_1, a_r)(a_1, a_{r-1}) \dots (a_1, a_2)$, thus (3) follows from (1). □

EXAMPLE 6.3. The cycle decomposition of

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

is $(1,6)(2)(3,5,4)$.

REMARK 6.4. For a cycle $(a_1, \dots, a_s)$ we call $\{a_1, \dots, a_s\}$ the support of the cycle. In the above theorem for any element $\sigma \in S_n$ we get a cycle decomposition $\sigma = \sigma_1 \dots \sigma_r$, such the the supports of the cycles σ_i are the orbits of $\langle \sigma \rangle$. In particular the supports of the σ_i give a decomposition of $\{1, \dots, n\}$ in to disjoint sets, but some of the cycles can be 1-cycles.

EXERCISE 6.5. Show that the cycle decompostion of an element $\sigma \in S_n$ is unique up to reordering: Let $\sigma = \sigma_1 \dots \sigma_r = \sigma'_1 \dots, \sigma'_s$ be two cycle decompostions such that the union of the supports of the σ_i is equal to the union of the supports of the σ'_i. Then $r = s$ and there is a bijection $f : \{1, \dots, r\} \to \{1, \dots, s\}$ with $\sigma_i = \sigma'_{f(i)}$ for all i.

Now we want to describe the conjugacy classes in S_n. We will see that the conjugacy class can be read off directly from the cycle decomposition of σ. In fact the conjugacy classes of S_n are parametrized by partitions of n.

Recall that a partition of a positive integer n is a tuple of positive integers $(n_1, n_2, \dots, n_r)$ with $n_1 \geq n_2 \geq n_2 \geq \dots \geq n_r$, such that $n_1 + \dots + n_r = n$. For instance the partitions of 4 are (4), $(3,1)$, $(2,2)$, $(2,1,1)$, $(1,1,1,1)$.

DEFINITION 6.6. Let $\sigma \in S_n$ and let $\sigma = \sigma_1 \dots \sigma_r$ be its cycle decomposition. By reordering the σ_i we assume that $l(\sigma_1) \geq l(\sigma_2) \geq \dots \geq l(\sigma_r)$. Then the tuple $(l(\sigma_1) \geq l(\sigma_2) \geq \dots \geq l(\sigma_r))$ is called the *cycle type* of σ. As the supports of the cycles are the orbits of σ on $\{1, \dots, n\}$, we find that $\sum_{i=1}^{r} l(\sigma_i) = n$. Thus the cycle type of σ is a partition of n.

LEMMA 6.7. *Two permutations $\sigma, \tau \in S_n$ are conjugated if and only if they have the same cycle type.*

PROOF. Let $\sigma = \sigma_1 \dots \sigma_r$ with $\sigma_i = (a_1^i, \dots, a_{m_i}^i)$. Let $\tau \in S_n$. For all $i = 1, \dots, r)$ and all $j = 1, \dots, m_i)$, we put $b_j^i = \tau(a_j^i)$. Then we have

$$\tau\sigma\tau^{-1}(b_j^i) = \tau\sigma(a_j^i) = \begin{cases} \tau(a_{j+1}^i) = b_{j+1}^i & j < m_i \\ \tau(a_1^i) = b_1^i & j = m_i \end{cases}.$$

Thus $\tau\sigma\tau^{-1} = \tau_1 \dots \tau_r$ with $\tau_i = (b_1^i, \dots, b_{m_i}^i)$. Therefore σ and $\tau\sigma\tau^{-1}$ have the same cycle type.

Conversely if σ and π have the same cycle type, we can write $\sigma = \sigma_1 \dots \sigma_r$ with $\sigma_i = (a_1^i, \dots, a_{m_i}^i)$ and $\pi = \pi_1 \dots \pi_r$ with $\sigma_i = (b_1^i, \dots, b_{m_i}^i)$ (the numbers r and m_i are the same). Then we define a bijection $\tau : \{1, \dots, n\} \to \{1, \dots, n\}$ by $\tau(a_j^i) := b_j^i$. By the argument above $\tau\sigma\tau^{-1} = \pi$. □

COROLLARY 6.8. *The number of conjugacy classes of S_n is equal to the number of partitions of n.*

PROOF. The cycle type defines an injective map

$$c : \{ \text{ conjugacy classes of } S_n \} \mapsto \{ \text{ partitions of } n \}.$$

We only have to see that this map is surjective; thus for every partition P of n we have to find a permutation σ whose cycle type is P. Let $P = (n_1, \dots, n_r)$, then we put

$$\sigma = (1, \dots, n_1)(n_1 + 1, \dots, n_1 + n_2) \dots (n_1 + \dots n_{r-1}, \dots, n).$$

□

For instance we see that S_4 has 5 conjugacy classes.

The sign of a permutation

An important invariant of a permutation σ is its sign, which can be defined as $(-1)^m$, where m is the number of transpositions used to write σ.

DEFINITION 6.9. Let $n \geq 2$ be a positive integer. For a permutation $\sigma \in S_n$, the *sign* of σ is $\varepsilon(\sigma) := \prod_{i>j} \frac{\sigma(i)-\sigma(j)}{i-j}$. The product is over all $i, j \in \{1, \dots, n\}$ with $i > j$.

LEMMA 6.10. (1) *$\varepsilon(\sigma) = (-1)^m$, where m is the number of pairs (i, j) is with $i > j$ and $\pi(i) < \pi(j)$.*
(2) *The map $\varepsilon : S_m \to (\{1, -1\}, \cdot)$ is a group homomorphism.*
(3) *If $\sigma = \tau_1 \dots \tau_k$ where the τ_i are transpositions, then $\varepsilon(\sigma) = k$.*

PROOF. (1)

$$\begin{aligned}\prod_{i>j}(\sigma(i) - \sigma(j)) &= \prod_{\substack{i>j \\ \sigma(i)>\sigma(j)}} (\sigma(i) - \sigma(j)) \cdot \prod_{\substack{i>j \\ \sigma(i)<\sigma(j)}} (\sigma(i) - \sigma(j)) \\ &= (-1)^m \prod_{\sigma(i)>\sigma(j)} (\sigma(i) - \sigma(j)) = (-1)^m \prod_{i>j}(i - j).\end{aligned}$$

Here the last equality is because the factors on the left hand side are just a permulation of the factors on the right hand side.

(2) For $\sigma, \tau \in S_n$ we have

$$\prod_{i>j} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{i - j} = \prod_{i>j} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)} \cdot \prod_{i>j} \frac{\tau(i) - \tau(j)}{i - j},$$

and

$$\prod_{i>j} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)} = \prod_{\tau(i)>\tau(j)} \frac{\sigma(\tau(i)) - \sigma(\tau(j))}{\tau(i) - \tau(j)} = \prod_{i>j} \frac{\sigma(i) - \sigma(j)}{i - j}.$$

Here the first equality is because whenever $i > j$ and $\tau(i) < \tau(j)$, then both numerator and denominator change sign. The second equality is just because the factors on the left and on the right hand side are just permutations of each other.

(3) It is clear that $\varepsilon((1,2)) = (-1)$. Thus, if $\tau = (i,j)$, and σ is a permutation with $\sigma(1) = i$, $\sigma(2) = j$, then we see that $\tau(1,2)\tau^{-1} = (i,j)$, thus $\varepsilon((i,j)) = \varepsilon(\tau)(-1)\varepsilon(\tau)^{-1} = (-1)$. Thus (3) follows from (2). □

DEFINITION 6.11. A permutations $\sigma \in S_n$ is called *even* if $\varepsilon(\sigma) = 1$, and *odd* if $\varepsilon(\sigma) = -1$.

DEFINITION 6.12. The set A_n of even permutations in S_n is a normal subgroup (because it is the kernel of the surjective group homomorphism $\varepsilon : S_n \to (\{1,-1\})$. It is called he *alternating group* of degree n.

7. Operations on Subsets

Let a group G operate on a set S. Then it also operates on the subsets of S. We want to make some observations about this action, mostly in case $S = G$. This will serve as a preparation for the proof of the Sylow theorems.

DEFINITION 7.1. Let $U \subset S$ be a subset. Then for $g \in G$ the set

$$gU := \{gu \mid u \in U\}$$

is a subset of S. It is easy to see that this defines an operation of G on the set of subsets of S by $g \cdot U := gU$.

We can restrict the operation to subsets U of S of a given order $|U|$. We know that multiplication by $g : U \to gU$ defines a bijection, therefore U and gU have the same order.

The *stabilizer* G_U of a subset U is the set of all $g \in G$ with $gU = U$. It is clear the G_U is a subgroup of G.

Note that that an element $g \in G$ is in the stabilizer G_U does not mean that $gu = u$ for all $u \in U$, but just that $gU = U$, in other words that $gu \in U$ for all $u \in U$.

PROPOSITION 7.2. *Let a group G act on a set S, and let U be a subset of S. Then $G = G_U$ if and only if U is a union of G orbits on S.*

PROOF. $G = G_U$ if and only if the orbit $G(u)$ of every $u \in U$ is contained in U. □

We want to consider two cases of this where S is the group G itself: The operation of G on itself by left translation, and the operation of a subgroup H of G by conjugation.

PROPOSITION 7.3. *Let G act on itself by left translation $g \cdot h = gh$. Let U be a subset of G. Then the order $|G_U|$ of the stabilizer of U divides the order $|U|$ of U.*

PROOF. Let $H = G_U$ be the stabilizer of U. Then, by the previous proposition, U is a union of orbits for the operation of H on U. Note that for any $x \in G$ the left multiplicaton $H \to G, h \mapsto hx$ is injective, with image the orbit $H(x)$. Thus $|U|$ is a multiple of $|H|$. □

Note that $|G_U|$ also divides $|G|$, thus if $|U|$ and $|G|$ have no common factor, then $G_U = \{1\}$.

Now we come to the operation by conjugation.

DEFINITION 7.4. Let G be a group, and let $H \subset G$ be a subgroup. For any $g \in G$ we have the conjugate subgroup

$$gHg^{-1} = \left\{ghg^{-1} \;\middle|\; h \in H\right\}.$$

It is obvious that gHg^{-1} is a subgroup of G, and the operation by conjugation defines an operation of G on the set of subgroups of G.

The stabilizer of the subgroup H for the operation of conjugation is called the *normalizer* of H, and denoted

$$N(H) = \left\{g \in G \;\middle|\; gHg^{-1} = H\right\}.$$

REMARK 7.5.
(1) By definition $N(H)$ is a subgroup of G and H is a subgroup of $N(H)$. Thus by Lagrange's Theorem $|H|$ divides $|N(H)|$, and $|N(H)|$ divides $|G|$.
(2) By definition the subgroup $H \subset G$ is a normal subgroup of G if and only if $gHg^{-1} = H$ for all $g \in G$, i.e. if and only if $N(H) = G$.
(3) Let c be the number of different conjugate subgroups of H. Then the orbit stabilizer theorem says. $|G| = |N(H)|c$.

8. The Sylow Theorems

The Sylow theorems describe the subgroups of $H \subset G$ of order $|H| = p^m$ a prime power in arbitrary finite groups. We have seen that if $H \subset G$ is a subgroup, then $|H|$ divides $|G|$. But the converse is in general not true, it is difficult to know for which divisors d of $|G|$ there exists a subgroup of order d. The first Sylow theorem will give a very partial converse: if p^m is the largest power of a prime p dividing $|G|$ there is a subgroup $H \subset G$ of order p^m. Such a group will be called a p-Sylow subgroup of $|G|$. The other Sylow theorems give more information about these p-Sylow subgroups. The Sylow theorems are the most advanced and difficult results we will prove in group theory, many of the results we proved before are used in this argument.

We will now state the Sylow theorems and give some applications. The proof of the Sylow theorems is at the end of this section.

In the following let G be a finite group, p a prime number and let always m be the largest nonnegative integer such that p^m divides $|G|$.

DEFINITION 8.1. A subgroup $H \subset G$ is called a p-Sylow subgroup, if $|H| = p^m$.

The first Sylow theorem is that p-Sylow subgroups always exist.

THEOREM 8.2. *(first Sylow theorem) There is a p-Sylow subgroup H of G.*

COROLLARY 8.3. *(Cauchy's Theorem) If a prime p divides $|G|$, then G contains an element of order p.*

PROOF. Let $H \subset G$ be a p-Sylow subgroup. Let $x \in H$ be an element different from 1. The order of x divides p^m, so it is p^r for some r with $0 < r \leq m$. Thus $x^{p^{r-1}}$ has order p. □

Some applications.

We review the product of groups from an example above and give a criterion for a group to a product of two groups.

DEFINITION 8.4. Let H, K be groups. The product $H \times K$ of H and K is

$$H \times K := \big\{(h,k) \;\big|\; h \in H, k \in K\big\},$$

with multiplication $(h_1,k_1)(h_2,k_2) = (h_1h_2, k_1k_2)$. The neutral element is $(1,1)$ and the inverse of (h,k) is (h^{-1},k^{-1}).

THEOREM 8.5. *Let H, K be subgroups of a group G. Assume*

(1) $G = HK = \big\{hk \;\big|\; h \in H, k \in K\big\}$,
(2) *H and K are normal subgroups of G,*
(3) $H \cap K = \{1\}$

Then $G \simeq H \times K$.

Condition (1) can also be replaced by (1') $|H||K| = |G|$.

PROOF. There is an obvious map $\theta : H \times K \to G; (h,k) \mapsto hk$. We want to show that it is an isomorphism. First we show it is a homomorphism: Let $h \in H, k \in K$. Then $h^{-1}k^{-1}hk = h^{-1}(k^{-1}hk) \in H$, because H is normal. and $h^{-1}k^{-1}hk = (h^{-1}k^{-1}h)k \in K$ because K is normal. By condition (3) $h^{-1}k^{-1}hk$, i.e. $hk = kh$, so every element of H commutes with every element of K. Now let (h_1,k_1), $(h_2,k_2) \in H \times K$. Then

$$\theta((h_1,k_1),(h_2,k_2)) = \theta((h_1h_2,k_1k_2)) = h_1h_2k_1k_2 = h_1k_1h_2k_2 = \theta((h_1,k_1))\theta((h_2,k_2)),$$

so θ is a homomorphism.

Now we shoe θ is injective. Let $(h,k) \in ker(\theta)$, i.e. $hk = 1$. Then $h = k^{-1}$ lies in $H \cap K = \{1\}$. Therefore $(h,k) = (1,1)$.

Finally we show θ is surjective: (1) just says θ is surjective. Under condition (1') we have θ is an injective map between two sets with the same number of elements, so it is surjective. □

We know that any cyclic group of order n is isomorphic to $\mathbb{Z}_n \simeq \mathbb{Z}/n\mathbb{Z}$.

COROLLARY 8.6. *If n and m are relatively prime, then $\mathbb{Z}_{nm} \simeq \mathbb{Z}_n \times \mathbb{Z}_m$.*

PROOF. We know by the Cauchy theorem (or also directly) that $\mathbb{Z}_{nm}$ has subgroups $H = \langle m \rangle$ and $K = \langle n \rangle$ of orders n and m. These are normail subgroups because $\mathbb{Z}_{nm}$ is commutative. Because n and m are relatively prime, we see that $H \cap K = \{0\}$. The result follows from the Theorem. □

COROLLARY 8.7. *Let p be a prime and G a group of order p^2. Then p is either isomorphic to $\mathbb{Z}_{p^2}$ or to $\mathbb{Z}_p \times \mathbb{Z}_p$.*

PROOF. By the previous Corollary G contains an element a of order p. Let $H = \langle a \rangle$. Let $b \in G \setminus H$, and $H' = \langle b \rangle$.

If $H' = G$, then G is cyclic.

Assume $H' \neq G$. Then $|H'|$ is a divisor of p^2, different from $1, p^2$, thus $|H'| = p$, and H' is cyclic. $H' \cap H$ is a subgroup of H and thus its order is a divisor of p. If $|H' \cap H| = p$, then $H = H'$, which is impossible because $b \notin H$. Thus $H \cap H' = \{1\}$. Thus by the theorem $G = H \times H'$. □

We recall the definition of the dihedral group from an earlier example. The dihedral group D_{2n} is a group of order $2n$ generated by 2 elements a, b with $a^n = 1$, $b^2 = 1$, $bab = a^{-1}$. It is easy to see that this determines D_n up to isomorphism, in fact we can use the relations to give a list of all elements of D_n:

$$D_n = \left\{ a^i b^j \;\middle|\; 0 \leq i < n,\ 0 \leq j < 2 \right\}.$$

THEOREM 8.8. *Let $p \geq 3$ be a prime number. Every group of order $2p$ is cyclic or dihedral.*

PROOF. Let G be a group of order $2p$. By Cauchy's theorem G contains an element a of order n and an element b of order 2. Let $H := \langle a \rangle$, $K := \langle b \rangle$. Then $[G : H] = 2$, so by an exercise H is normal in G. It is clear that $G = \langle a, b \rangle$. Therefore we have $bab = bab^{-1} = a^i$ for some $i \in \mathbb{Z}$. Thus

$$a = b^2 a b^2 = b(bab)b = ba^i b = (bab)^i = (a^i)^i = a^{i^2}.$$

In other words $a^{i^2-1} = 1$. Thus we get p divides $i^2 - 1$, and therefore, as p is a prime, p divides $i - 1$ or p divides $i + 1$.

If p divides $i - 1$, then $a^{i-1} = 1$, i.e. $a^i = a$. In other words $bab = a$, $ab = ba$. As G is generated by a and b, it follows that G is abelian. So $\langle b \rangle$ is also normal in G, and therefore $G = \langle a \rangle \times \langle b \rangle \simeq \mathbb{Z}_p \times \mathbb{Z}_2 \simeq \mathbb{Z}_{2p}$. Thus G is cyclic.

If p divides $i + 1$, then $a^{i+1} = 1$, i.e. $a^i = a^{-1}$, $bab = a^{-1}$. Thus G is isomorphic to D_p. □

The second Sylow theorem says in particular that all p-Sylow subgroups are conjugated.

THEOREM 8.9. *(Second Sylow theorem) Let K be a subgroup of G whose order is divisible by p, and let H be a p-Sylow subgroup of G. Then there is a conjugate group $H' = gHg^{-1}$, such that $K \cap H'$ is a p-Sylow group of K.*

COROLLARY 8.10.
(1) *Let K be a subgroup of G which is a p-group, then K is contained in a Sylow p-subgroup of G,*
(2) *The Sylow p-subgroups of G are all conjugate.*

PROOF. The conjugation $H \to gHg^{-1}, h \mapsto ghg^{-1}$ is a bijection, thus a conjugate of a Sylow subgroup is a Sylow subgroup.

(1) By definition if K is a p-subgroup of G, then K is a Sylow subgroup of itself. Thus by the Theorem, if H is a Sylow subgroup of G, then there exists a conjugate $H' = gHg^{-1}$, such that $H' \cap K = K$, i.e. such that $K \subset H'$. Thus H' is a Sylow subgroup containing K.

(2) Let H, K be Sylow subgroups. Then there exists a conjugate H' of H containing K. As H' and K have the same order, they are equal. □

The third Sylow theorem tells us something about how many p-Sylow subgroups there are.

THEOREM 8.11. *(Third Sylow Theorem). Let $|G| = n = p^m r$, such that p does not divide r. Let s be the number of p-Sylow subgroups of G. Then s divides r, and s is congruent 1 modulo p, i.e. $s = ap + 1$ for some integer $a \geq 0$.*

We give one further application to the classification of finite groups.

THEOREM 8.12. *Let G be a group of order $|G| = pq$, where p, q are prime numbers with $p > q$ and q does not divide $p - 1$. Then G is cyclic.*

PROOF. Let N_p be the number of p-Sylow subgroups of G. Then, by the third Sylow theorem, $N_p \equiv 1 \mod p$ and N_p divides q. Therefore we get $N_p = 1$, because $p > q$. Let $H \subset G$ be the unique p-Sylow subgroup of G, Any conjugate gHg^{-1} of H is also a p-Sylow subgroup, therefore $gHg^{-1} = H$, i.e. H is a normal subgroup.

Let N_q be the number of q-Sylow subgroups of G. Then $N_q \equiv 1 \mod q$ and N_q divides p. Therefore we have $N_q = 1$ or $N_q = p$ because p is a prime number. If $N_q = p$, we have that $p \equiv 1 \mod q$, so q divide $p - 1$, a contradiction. Therefore $N_q = 1$, and by the same argument as above the unique q-Sylow subgroup K of G is a normal subgroup. We know H and K are cyclic of orders p and q, so clearly $H \cap K = \{1\}$. Thus the theorem above applies ad $G = H \times K = \mathbb{Z}_p \times \mathbb{Z}_q = \mathbb{Z}_{pq}$. □

Now we will prove the Sylow theorems.

The proof of the first Sylow theorem is truely remarkable.

PROOF. *(of the first Sylow theorem)*

We start with an elementary lemma.

LEMMA 8.13. *The number of subsets of order p^m of a set with $n = p^m r$ elements (where p does not divide r) is*

$$N := \binom{n}{p^m} = \frac{n(n-1)\cdots(n-k)\cdots(n-p^m+1)}{p^m(p^m-1)\cdots(p^m-k)\cdots 1}.$$

Furthermore p does not divide N.

PROOF. It is well-known that the number of subsets of order p^m is this binomial coefficient. To see that N is not divisible by p, note that whenever p^l divides a factor $(n-k)$ in the numerator the same factor p^l also divides (p^m-k). Write $k=p^i s$ with p not dividing s, then $i<m$. Therefore both $(n-k)$ and p^m-k are both divisible by p^i and not by p^{i+1}. □

Now let S be the set of all subsets M of G of order p^m. We decompose S into orbits for the left multiplication $g\cdot M=gM$. Thus

$$N=|S|=\sum_{\text{orbits } O}|O|.$$

As p does not divide N, there is an orbit $O=G(M)$ whose order is not divisible by p. By the Proposition **??** the order of the stabilizer G_M is a power of p. By the orbit stabilizer theorem, we have $p^m r=|G|=|G_M|\cdot|G(M)|$. As $G(M)$ is not divisible by p we have $|G_M|=p^m$. G_M is the required p-Sylow subgroup. □

PROOF. *(of the second Sylow theorem)* Let K be a subgroup of G and H a p-Sylow subgroup. We have to show that for a conjugate subgroup H' to H, the intersection $K\cap H'$ is a Sylow subgroup of K.

Now let $S=G/H$ be the set of left cosets. Recall that G acts transitively on these cosets and that H is the stabilizer of $s=1H$. We see from this that the stabilizer of as is aHa^{-1}.

Restrict the operation of G on S to K and decompose S into K-orbits. As H is a p-Sylow group, the order S is prime to p. Therefore there is a K-orbit $O=K(as)$ on S whose order is prime to p. Let $H'=aHa^{-1}$. This is the stabilizer of as, for the operation of G. Therefore the stabilizer of the restriction of the operation to K is $H'\cap K$, and the index $[K:H'\cap K]$ is $|O|$, which is prime to p. Since H' is conjugate to H it is a p-group. Thus $H'\cap K$ is a p-group. Thus $H'\cap K$ is a Sylow subgroup of K. □

PROOF. *(of the third Sylow theorem)* By Corollary **??**, the Sylow subgroups of G are all conjugate to one Sylow subgroup H. Thus the number of Sylow subgroups is $s=[G:N]$, where N is the normalizer of H. Since H is a subgroup of N we get that $[G:N]$ divides $[G:H]=r$.

To show that $s\equiv 1 \mod p$, decompose the set $\{H_1,\dots,H_s\}$ of Sylow subgroups into orbits for the operation of conjugation by $H=H_1$. An orbit consists of a single group H_i if and only if H is contained in the normalizer N_i of H_i.

If this is the case, then H and H_i are both Sylow subgroups of N_i and H_i is normal in N_i. By Corollary **??**, we get $H=H_i$. Thus there is only one H orbit of order 1, namely $\{H\}$. The other orbits have orders divisible by p because their orders divide $|H|$, by the orbit stabilizer theorem. This shows $s\equiv 1 \mod p$. □

9. Exercises

(1) Find a subgroup H of S_3 of order 3, i.e. with $|H| = 3$.
(2) Let G be a group. The center $C(G)$ of G is

$$C(G) := \{g \in G \mid ga = ag,\ \forall_{a\in G}\},$$

   (a) Show $C(G)$ is a subgroup of G,
   (b) Show $C(G)$ is commutative.
   (c) What is the center of S_3.
(3) Prove that every subgroup of a cyclic group is cyclic.
(4) Let G be a finite group. Let $K \subset H$ and $H \subset G$ be subgroups. Show that $[G : K] = [G : H][H : K]$.
(5) Let G be a group and H, K normal subgroups. Then $H \cap K$ is a normal subgroup of G.
(6) Let G be a group and H a subgroup of index 2. Show H is a normal subgroup of G.
(7) Let G be a group. Let G' be the subgroup of G generated by

$$U = \{xyx^{-1}y{-}1 \mid x, y \in G\}.$$

   (a) Show G' is a normal subgroup of G.
   (b) Show G/G' is abelian.
(8) Let G be a group. Show that $inn(G)$ is a normal subgroup of $Aut(G)$.
(9) Give and example of a group G and subgroups $K \subset H \subset G$, such that K is a normal subgroup of H and H is a normal subgroup of G, but K is not a normal subgroup of G.
(10) Let H, K be normal subgroups of a group G with $H \cap K = \{1\}$. Show that $hk = kh$ for all $h \in H$, $k \in K$.
(11) Let

$$\sigma := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 4 & 5 & 1 & 6 & 7 & 9 & 8 \end{pmatrix} \in S_9.$$

Find the cycle decomposition of σ, find the cycle type of σ, and write an element of S_9 which is conjugated to σ.
(12) In the following verify if the maps $\varphi : G \to G$ are homomorphisms of groups. If they are, determine their kernel.
   (a) $G = (\mathbb{R} \setminus \{0\}, \cdot)$ the nonzero real numbers with multiplication. $\varphi(x) = x^2$.
   (b) $G = (\mathbb{R}, +)$ the real numbers with addition. $\varphi(x) = x + 1$.
   (c) G and abelian group. $\varphi(x) = x^3$.
(13) Let G be a group, let $\varphi : G \to G$ be an automorphism, let $Z(G)$ be the center of G. Show $\varphi(Z(G)) \subset Z(G)$.
(14) Let G be a cyclic group of prime order. Determine $Aut(G)$.
(15) Let G be a group, H a subgroup. . If $[G : H] = 2$, then H is normal in G.

(16) If $G/Z(G)$ is cyclic, then G is abelian.
(17) An r-cycle is even if and only if r is odd.
(18) Let G be a group of order 30.
  (a) Show a 3-Sylow or a 5-Sylow subgroup is normal in G.
  (b) Show every 3-Sylow and every 5-Sylow subgroup is normal in G.
  (c) Show G has a normal subgroup of order 15.
(19) If G is a group of order 385, show that its 11-Sylow subgroup is normal and its 7-Sylow subgroup is in the center of G.
(20) Let G be a group of order 231. Show the 11-Sylow subgroup is in the center of G.
(21) If G is a group of order p^2q with p and q primes, show that G has a nontrivial normal subgroup.
(22) How many elements of order 5 are contained in a group of order 20.
(23) Prove: No group of order pq with p and q prime numbers, is simple.
(24) Let G be a group of order $p^m r$. Show G contains a subgroup of order p^e for every $e \leq m$.
(25) Classify the groups of order 33.

CHAPTER 2

Rings

1. Definition of a ring, examples and first properties

From high school and even from earlier, everybody is aware of the integers. Integers can be added, subtracted and multiplied, and there is the distributive law relating addition and multiplication. This is formalized in the notion of a ring. Basically the axioms of a ring will guarantee that in a ring one can compute more or less like with integers. When dealing with rings we always want to keep the integers in mind as a basic example.

DEFINITION 1.1. A *ring* R is a nonempty set R together with two binary operations

$$\begin{aligned} + &: R \times R \to R \qquad \text{(addition)} \\ \cdot &: R \times R \to R \qquad \text{(multiplication)} \end{aligned}$$

and a distinguished element $0 \in R$, such that

(1) $(R, +)$ is an abelian group with neutral element 0.
(2) $\cdot$ is associative

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \quad \text{for all } a, b, c \in R.$$

(3) We have the distributive laws

$$\begin{aligned} a \cdot (b + c) &= a \cdot b + a \cdot c, \\ (b + c) \cdot a &= b \cdot a + c \cdot a. \end{aligned}$$

We call $(R, +)$ the *additive group* of R. Usually we write ab for $a \cdot b$.

EXAMPLE 1.2. The integers $\mathbb{Z}$ are a ring with the usual addition and multiplication.

REMARK 1.3. The distributive laws can also be reformulated as follows: For $a \in R$ let $l_a : R \to R, b \mapsto ab$ the multiplication by a from the left, and let $r_a : R \to R, b \mapsto ba$ the multiplication by a on the right. Then the distributive laws say precisely that l_a and r_a are group homomorphisms $(R, +) \to (R, +)$.

A subset of a ring is a subring if, with the restriction of $+$ and $\cdot$, it is itself a ring.

DEFINITION 1.4. Let R be a ring. A subset $A \subset R$ is called a *subring* of R if

(1) A is a subgroup of the additive group of R.
(2) For all $a, b \in A$ we have $ab \in A$.

(Equivalently $0 \in A$ and $a + b$, $-a$, $ab \in A$ for all $a, b \in A$).

EXAMPLE 1.5. Let

$$2\mathbb{Z} := \{2n \mid n \in \mathbb{Z}\}$$

be the set of even integers. Then $2\mathbb{Z}$ is a subring of $\mathbb{Z}$. Note that the set of odd integers is not a subring of $\mathbb{Z}$.

The integers have some further useful properties:

(1) $ab = ba$ for all $a, b \in \mathbb{Z}$,
(2) $1a = a1 = a$ for all $a \in \mathbb{Z}$.

DEFINITION 1.6. A ring R is called *commutative* if

$$ab = ba \quad \text{for all } a, b \in R$$

i.e. if the multiplication is commutative. (Note that, by the definition of a ring, the addition is already commutative).

DEFINITION 1.7. An element $1 \in R \setminus \{0\}$ is called a *unit element* of R if

$$a \cdot 1 = 1 \cdot a = a \quad \text{for all } a \in R.$$

Note that we are assuming that $1 \neq 0$. (Note also that 1 is then unique). If R contains a unit element, it is called a unital ring or a ring with 1.

In this course we will mostly be interested in commutative rings with 1. After the first few lectures we will restrict our attention to them.

EXAMPLE 1.8.
(1) The integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$ with the usual addition and multiplication are all commutative rings with 1. We briefly recall the complex numbers. $\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R}\}$ where the ring structure is given by

$$(a + ib) + (c + id) = a + c + i(b + d),$$
$$(a + ib)(c + id) = (ac - bd) + i(ad + bc),$$

i.e. one computes normally, using that $i^2 = (-1)$.
(2) Let $\mathbb{Z}[i] := \{n + im \in \mathbb{C} \mid n, m \in \mathbb{Z}\}$. $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. $\mathbb{Z}[i]$ is called the ring of *Gaussian integers*.
(3) $2\mathbb{Z}$ is a commutative ring (as it is a subring of $\mathbb{Z}$, but it is not a ring with 1).
(4) Let $M_{n\times n}(\mathbb{R})$ be the set of $n \times n$ matrices with entries in $\mathbb{R}$, with the usual addition and multiplication of matrices. Then $M_{n\times n}(\mathbb{R})$ is a ring with 1. The unit element is the diagonal matrix with 1 on the diagonal. But $M_{n\times n}(\mathbb{R})$ is not commutative for $n \geq 2$. $M_{n\times n}(\mathbb{R})$ is the most common example of a noncommutative ring. If we have to think of noncommutative rings we always first look at matrices.

(5) Let $\mathbb{C}^\infty(\mathbb{R})$ be the set of $\mathbb{C}^\infty$ functions $f : \mathbb{R} \to \mathbb{R}$. Then $\mathbb{C}^\infty(\mathbb{R})$ is a commutative ring with 1 with pointwise addition and multiplication:

$$(f+g)(x) = f(x) + g(x), \qquad (fg)(x) = f(x)g(x).$$

The zero element is the constant function 0 and the unit element is the constant function 1.

(6) Let $k \in \mathbb{Z}$ and let $\mathbb{Z}_k$ be the set of integers mod k. The elements of $\mathbb{Z}_k$ are $0, 1, \ldots, k-1$, and addition and multiplication are defined by

$$\underline{n} + \underline{m} = \underline{n+m}, \quad \underline{n}\,\underline{m} = \underline{nm}$$

where $\underline{n}$ denotes the rest of n after division by k. Check that $\mathbb{Z}_k$ is a commutative ring with zero element 0 and unit element 1.

We can do computations in a ring in much the same way as in the integers. The basic rules carry over.

REMARK 1.9. Let R be ring. Then for all $a, b \in R$

(1) $a0 = 0a = 0$,
(2) $(-a)b = a(-b) = -(ab)$,
(3) $(-a)(-b) = ab$,
(4) if R is unital, then $(-1)a = a(-1) = -a$.

PROOF. These are straightforward verifications. (1) $a0 = a(0+0) = (a0+a0)$. As $(R, +)$ is a group, we can subtract $a0$ from both sides to get $a0 = 0$. (2) $ab + (-a)b = (a+(-a))b = 0b = 0$. (3) follows by applying (2) twice. (4) $(-1)a = -(1a) = -a$. □

In future we will write $a - b$ for $a + (-b)$.

We come to some further properties of rings. In the integers $\mathbb{Z}$ we know that the only way how a product ab can be 0 is that $a = 0$ or $b = 0$. This is a very useful property of the integers. It implies that we can cancel factors: if $ab = ac$ with $a \neq 0$, then $b = c$.

DEFINITION 1.10. Let R be a commutative ring. An element $a \in R \setminus \{0\}$ is called a *zero divisor* if there is an element $b \in R \setminus \{0\}$ with $ab = 0$. A commutative ring R with 1, which contains no zero divisors is called an *integral domain*.

REMARK 1.11. (Cancellation) Let R an integral domain, then the cancellation law is valid, i.e., if $ab = ac$ and $a \neq 0$, then $b = c$.

PROOF. Let $ab = ac$ and $a \neq 0$. Then $a(b-c) = 0$. As R contains no zero divisors it follows that $(b - c) = 0$. □

Integral domains are particularly nice rings. Most of the rings we will study are integral domains.

EXAMPLE 1.12. (1) $\mathbb{Z}$ is an integral domain.
(2) In $\mathbb{Z}_6$ we see that $\underline{2}\,\underline{3} = \underline{0}$. Thus $\mathbb{Z}_6$ is not an integral domain.

(3) Let $p \in \mathbb{Z}_{>0}$ be a prime number. If n and m are not divisible by p, then also nm is not divisible by p. Thus $\mathbb{Z}_p$ is an integral domain.

DEFINITION 1.13. Let R be a ring with 1. An element $a \in R$ is called a *unit* of R if it has a multiplicative inverse, i.e. if there exists an element $b \in R$ with $ab = ba = 1$. The set of units in R is denoted by R^*. Check as an exercise that if a is a unit, then there is a unique element $b \in R$ with $ab = ba = 1$. We write $b := a^{-1}$.

EXAMPLE 1.14. (1) $\mathbb{Z}^* = \{1, (-1)\}$.
(2) $\mathbb{Z}[i]^* = \{1, (-1), i, -i\}$.
(3) $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$.
(4) The units in $M_{n\times n}(\mathbb{R})$ or $M_{n\times n}(\mathbb{C})$ are the matrices with nonzero determinant.

We see that in each of the above $(R^*, \cdot)$ is a group. This is always true:

PROPOSITION 1.15. *Let R be a ring with 1. Then $(R^*, \cdot)$ is a group.*

PROOF. This is easy. We know that the multiplication is associative. We know that $1 \in R^*$ is the neutral element for the multiplication, and that for every $a \in R^*$ there is an element $b \in R^*$ with $ab = ba = 1$, i.e. an inverse element. We only need to show that for $a, b \in R^*$ also $ab \in R^*$. Let $c, d \in R$ with $ac = ca = 1$, $bd = db = 1$. Then $(ab)(dc) = ac = 1$ and $(dc)(ab) = db = 1$. Thus R^* is a group. □

DEFINITION 1.16. $(R^*, \cdot)$ is called the *multiplicative group* of R.

NOTATION 1.17. Let R be a ring, $a \in R$ and $n \in \mathbb{Z}_{>0}$. Then we write

$$na := \underbrace{a + \ldots + a}_{n \text{ times}},$$
$$a^n := \underbrace{a \cdot \ldots \cdot a}_{n \text{ times}}.$$

We also write $0a := a$ and if R is a ring with 1 we write $a^0 := 1$. Finally we write $(-n)a := -(na)$, and, if a is a unit, $a^{-n} := (a^n)^{-1}$. It is easy to check (and left as an exercise) that the usual rules apply for $n, m \in \mathbb{Z}$

$$(n + m)a = na + ma, \quad (nm)a = n(m(a)),$$
$$a^n a^m = a^{n+m}, \quad (a^n)^m = a^{nm},$$

whenever both sides of the equation make sense.

Polynomial rings. We come now to the most important examples of rings. These are the rings $R[x]$ of polynomials with coefficients in a ring R. Already from high school we are familiar with polynomials $f = \sum_{i=0}^{n} a_i x^i$ with coefficients in the real numbers $\mathbb{R}$. One learns how they can be added, subtracted and multiplied in the usual way. We need to give a precise definition.

DEFINITION 1.18. Let R be a ring with 1. A *polynomial* f in x with coefficients in R is a formal expression

$$f = a_0 + a_1 x^1 + \ldots + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

with $a_i \in R$ and $n \in \mathbb{Z}_{\geq 0}$. We call a_i the *coefficient* of x^i of f. Two polynomials $f = \sum_{i=0}^{n} a_i x^i$, $\sum_{j=0}^{m} b_j x^j$ with $n \leq m$ are considered equal (and we write $f = g$), if $b_j = 0$ for $j > n$ and $b_i = a_i$ for $i = 0, \ldots, n$. Thus if $f = \sum_{i=0}^{n} a_i x^i$, $\sum_{j=0}^{m} b_j x^j$ are two polynomials, we can always assume that $n = m$. We denote by $R[x]$ the set of polynomials in x with coefficients in R. The polynomial $a_0 = a_0 x^0$ is identified with the element $a_0 \in R$. A polynomial of this form is called a *constant polynomial*. Thus R is the subset of $R[x]$ of constant polynomials.

REMARK 1.19. For a polynomial $f = \sum_{i=0}^{n} a_n x^n \in R[x]$ we can define a function

$$R \to R;\ b \mapsto f(b) = \sum_{i=0}^{n} a_n b^n.$$

Note however that differently from analysis for us a polynomial is not this function, but just a formal expression.

DEFINITION 1.20. We define a ring structure on $R[x]$. If $f = \sum_{i=0}^{n} a_i x^i$ and $g = \sum_{i=0}^{n} b_i x^i$ then we define

$$f + g := \sum_{i=0}^{n} (a_i + b_i) x^i,$$

$$fg := \sum_{k=0}^{2n} \Big(\sum_{i+j=k} a_i b_j \Big) x^k.$$

We see that with this definition the addition is defined as usual by adding the corresponding coefficients. The multiplication is as usual given by putting $(ax^i)(bx^j) = abx^{i+j}$ and formally applying the distributive law.

It is easy to check (and left as an exercise) that with these operations $R[x]$ is a ring with 1. The zero element is $0 \in R$ and the unit element is $1 \in R$. If R is commutative, then also $R[x]$ is commutative. We also see that R is a subring of $R[x]$. We call $R[x]$ the *polynomial ring* in x over R.

If some coefficients of a polynomial are zero we usually do not write them. Thus we write $x^3 + 2x$ for $1x^3 + 0x^2 + 2x + 0x^0$.

DEFINITION 1.21. Let R be a commutative ring with 1, and let $f \in R[x] \setminus \{0\}$. Write $f = \sum_{i=0}^{n} a_i x^i$ with $a_n \neq 0$. Then n is called the *degree* of f and denoted by $deg(f)$. a_n is called the *leading coefficient* of f.

If R is an integral domain, then also $R[x]$ is an integral domain and its units are just the units of R.

REMARK 1.22. Let R be an integral domain, $f, g \in R[x] \setminus \{0\}$.

(1) $fg \neq 0$ and $deg(fg) = deg(f) + deg(g)$.
(2) $R[x]$ is an integral domain.
(3) $R[x]^* = R^*$, the units in $R[x]$ are just the units in R.

PROOF. (1) Let $f = \sum_{i=0}^{m} a_i x^i$ with $a_m \neq 0$, $g = \sum_{i=0}^{n} b_i x^i$ with $b_n \neq 0$, then

$$fg = \sum_{i=0}^{n+m} c_i x^i, \qquad c_i = \sum_{k+l=i} a_k b_l,$$

in particular $c_{n+m} = a_m b_n \neq 0$. Thus $fg \neq 0$ and $deg(fg) = n + m$.

(2) follows from (1).

(3) It is clear that every unit of R is a unit of $R[x]$ (take as its inverse just its inverse in R). Let $f \in R[x]^*$. Then there exists $g \in R[x]$ with $fg = 1$. By (1) this implies $deg(f) + deg(g) = 0$, i.e. $deg(f) = 0$. Thus $f \in R$ and thus $f \in R^*$. □

Finally we come to the definition of a division ring and a field. Note that 0 is never a unit in a ring R (because $0a = a0 = 0$ for all a). A ring with 1 in which every nonzero element is a unit is called a division ring.

DEFINITION 1.23. A ring R with 1 is called a *division ring*, if every nonzero element is a unit, i.e. $R^* = R \setminus \{0\}$.

The most important case are the commutative division rings: the fields.

DEFINITION 1.24. A commutative division ring R is called a *field*. Explicitly this means the following. A set R with two binary operations $+$, $\cdot$ and two distinguished elements $0 \neq 1$ is called a field if

(1) $(R, +)$ is a commutative group with neutral element 0,
(2) $(R \setminus \{0\}, \cdot)$ is a commutative group with neutral element 1,
(3) the distributive law holds:

$$a(b + c) = ab + ac \quad \text{for all } a, b, c \in R.$$

The concept of field is of fundamental importance. While the notion of a ring formalizes that one is able to compute like in the integers, the notion of a field formalizes that one can compute like in the rational numbers. The second half of this course will be devoted to the theory of fields.

EXAMPLE 1.25. $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are fields. In $\mathbb{C}$ the inverse to $a + bi$, $a, b \in \mathbb{R}$ is $(a - bi)/(a^2 + b^2)$.

We want to give one example of a division ring which is not a field, the Quaternions.

EXAMPLE 1.26. (Quaternions) The most famous example of a noncommutative division ring are the Quaternions. Recall that in $\mathbb{C}$ we have *complex conjugation* i.e. for $c = a + ib$ $(a, b \in \mathbb{R})$ we have $\overline{c} = a - ib$, which fulfils the following properties: $\overline{\overline{c}} = c$, $c\overline{c} = a^2 + b^2$, which is a positive real number unless $c = 0$, $\overline{cd} = \overline{c}\overline{d}$, $\overline{c + d} = \overline{c} + \overline{d}$. Let $Q := \left\{ \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} \;\middle|\; a, b \in \mathbb{C} \right\}$. Q is called the ring of quaternions. It is easy to check that Q is a subring of $M_{2\times 2}(\mathbb{C})$. We have $det \begin{pmatrix} a & b \\ -\overline{b} & \overline{a} \end{pmatrix} = a\overline{a} + b\overline{b}$, thus if the element is nonzero its inverse is given by $\frac{1}{a\overline{a}+b\overline{b}} \begin{pmatrix} \overline{a} & -b \\ \overline{b} & a \end{pmatrix}$. Thus Q is a division ring.

On the other hand $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} = - \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$. Thus Q is not a field.

We want to give one less standard example of a field, in particular we want to see that there are also fields with finitely many elements.

PROPOSITION 1.27. *Every finite integral domain is a field.*

PROOF. By the above we have to show that $R^* = R \setminus \{0\}$, that is we have to find an inverse for any nonzero $a \in R$. Recall that the distributive law $a(b + c) = ab + ac$ says that the map

$$\cdot a : (R, +) \to (R, +), b \mapsto ba$$

of multiplication by a is a group homomorphism. As R has no zero divisors, the kernel of $\cdot a$ is 0. Thus $\cdot a$ is injective. But it is one of the fundamental properties of finite sets, that an injective map of a finite set onto itself is also bijective. So let $c \in R$ be an element with $\cdot a(c) = 1$. Then $ac = 1$, thus a is a unit. □

EXAMPLE 1.28. In particular $\mathbb{Z}_p$ is a field for p a prime number. It is called the finite field with p elements and denoted F_p. Later we shall see that there are finite fields F_{p^n} with p^n elements for every $n > 0$.

Exercises

(1) Let R be a ring with 1, and let $a \in R$ a unit. Show that there is a *unique* $b \in R$ with $ab = ba = 1$.
(2) Show that the units in $\mathbb{Z}[i]$ are $\{1, -1, i, -i\}$.
(3) Show: The odd integers with usual addition and multiplication are not a ring.
(4) Let R be a ring and $a, b \in R$. Show $(a + b)^2 = a^2 + ab + ba + b^2$ (using only the definitions).
(5) An element a in a ring R is called nilpotent if there is an $n \in \mathbb{Z}_{>0}$ with $a^n = 0$. Show: if a is nilpotent, then $1 + a$ is a unit.

(6) Show: In a commutative ring R, $a+b$ is nilpotent if a and b are nilpotent. Show that this may be false if R is not commutative.
(7) Show that a commutative ring R is an integral domain if and only if for all $a, b, c \in R$ with $a \neq 0$, the relation $ab = ac$ implies $b = c$.
(8) Let U be a set, and let R be the set of subsets of U. Addition and multiplication on R are defined by

$$A + B := A \cup B, \quad A \cdot B := A \cap B.$$

Prove or disprove: R a ring.
(9) Let U be a set, and let R be the set of subsets of U. Addition and multiplication on R are defined by

$$A + B := (A \setminus B) \cup (B \setminus A), \quad A \cdot B := A \cap B.$$

Prove or disprove: R a ring.
(10) Let $GL(2, \mathbb{R}) := \{A \in M_{2\times 2}(\mathbb{R}) \mid det(A) \neq 0\}$. Is $GL(2, \mathbb{R})$ a subring of $M_{2\times 2}(\mathbb{R})$?
(11) Let R be a ring with $a^2 = a$ for all $a \in R$. Show: R is commutative. Hint: Show that $ab = -ba$ and that $a = -a$ for all $a, b \in R$.
(12) Determine the group of units in $\mathbb{Z}_{12}$.
(13) Let R be a ring and let $Z(R) := \{x \in R \mid xy = yx \text{ for all } y \in R\}$. Show that $Z(R)$ is a subring of R. $Z(R)$ is usually called the center of R.
(14) If R is a division ring, show that $Z(R)$ is a field.
(15) Let R be an integral domain and $a, b \in R$. Assume $a^n = b^n$ and $a^m = b^m$ for two relatively prime integers n, m. Show that $a = b$.
(16) Let m be an integer which is not divisible by the square of an integer > 1.
  (a) Show that $\mathbb{Z}[\sqrt{m}] := \{a + b\sqrt{m} \mid a, b \in \mathbb{Z}\}$ is an integral domain.
  (b) Find the units in $\mathbb{Z}[\sqrt{m}]$ for all $m < 0$.
(17) Show: A finite ring with 1 (not necessarily commutative) is a division ring.
(18) Let R be a commutative ring with 1. Determine the units in $R[x]$.

2. Homomorphisms, Ideals and Quotient Rings

When one studies sets with a structure, like groups and rings, the most important thing to look at are the maps between them which are compatible with the structure. For groups G, H these are the group homomorphisms, i.e. the maps $\varphi : G \to H$ which are compatible with the multiplication in the groups $\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2)$. For rings we have two operations $+$ and $\cdot$, so a ring homomorphism has to be compatible with both of them.

Definition 2.1. Let A, B be rings. A map $\varphi : A \to B$ is called a *ring homomorphism* if for all $a, b \in A$

(1) $\varphi(a + b) = \varphi(a) + \varphi(b)$,
(2) $\varphi(ab) = \varphi(a)\varphi(b)$.

The *image* of φ is $\varphi(A)$ and the *kernel* is $ker(\varphi) := \varphi^{-1}(0)$. A bijective ring homomorphism $\varphi : A \to B$ is called an *isomorphism*, in this case we say that A and B are *isomorphic*. A ring isomorphism $\varphi : A \to A$ is called an *automorphism*.

REMARK 2.2. (1) The identity map $id : R \to R; a \mapsto a$ is a ring isomorphism.
(2) If $\varphi : A \to B$ is a is an isomorphism, then also the inverse map φ^{-1} is a ring isomorphism.
(3) The composition of ring homomorphisms $\varphi : A \to B$, $\psi : B \to C$ is a ring homomorphism $\psi \circ \varphi : A \to C$.

REMARK 2.3. A ring homomorphism $\varphi : A \to B$ is in particular a homomorphism from the additive group of A to that of B. Thus results about group homomorphisms apply to ring homomorphisms. In particular φ is injective if and only of $ker(\varphi) = 0$.

EXAMPLE 2.4. The residue map $\mathbb{Z} \to \mathbb{Z}_m, n \mapsto \underline{n}$ sending n to the rest after division by m is a surjective ring homomorphism. The kernel is $m\mathbb{Z}$.

REMARK 2.5. Let $\varphi : A \to B$ be a ring homomorphism. Then $\varphi(A)$ is a subring of B.

PROOF. As φ is a homomorphism for the additive groups, we know that $\varphi(A)$ is a subgroup of the additive group of B. If $x = \varphi(a)$, $y = \varphi(b)$ are elements of $\varphi(A)$, then $xy = \varphi(a)\varphi(b) = \varphi(ab) \in \varphi(A)$. □

The kernel of a homomorphism of groups $\varphi : A \to B$ is not just a subgroup of A, but has a stronger structure: it is a normal subgroup. Also for homomorphisms of rings the kernel is not just a subring but something better; it is an ideal.

DEFINITION 2.6. A subset I of a ring R is called an *ideal*, if I is a subgroup of the additive group of R and for each $a \in I$ and all $x \in R$, we have $xa \in I$ and $ax \in I$. In particular ideals are subrings.

It is easy to see that 0, R are ideals in R and that the intersection of a set of ideals in R is an ideal in R.

DEFINITION 2.7. Let R be a commutative ring and let $a_1, \ldots, a_n \in R$. The *ideal generated by* $a_1, \ldots, a_n$ is the set

$$\langle a_1, \ldots, a_n \rangle := \{a_1 r_1 + \ldots + a_n r_n \mid r_1, \ldots, r_n \in R\}.$$

It is easy to check that this is an ideal. In particular $\langle a \rangle = aR = \{ar \mid r \in R\}$ for $a \in R$ is called a *principal ideal*. E.g. $\langle n \rangle = n\mathbb{Z}$ is an ideal in $\mathbb{Z}$.

REMARK 2.8. Let $a, b \in R$, then $\langle a \rangle = \langle b \rangle$ if and only if there is a unit u with $b = ua$.

PROOF. If $\langle a \rangle = \langle b \rangle$, then $b = va$ and $a = wb$ for elements $v, w \in R$. Thus $b = vwb$, i.e. $vw = 1$ and v is a unit. Conversely, if $b = ua$ for a unit u, then $b \in \langle a \rangle$ and thus $\langle b \rangle \subset \langle a \rangle$. On the other hand $a = u^{-1}b \in \langle b \rangle$ and thus $\langle a \rangle \subset \langle b \rangle$. □

Ideals are more important then subrings. We shall soon show that every ideal is the kernel of some ring homomorphism. Thus *the ideals are precisely the kernels of ring homomorphisms*. We know an analogous statement for normal subgroups: the normal subgroups are precisely the kernels of group homomorphisms. Thus ideals in rings are the analogues of normal subgroups in groups.

LEMMA 2.9. *Let $\varphi : A \to B$ be a homomorphism of rings.*

(1) *$ker(\varphi)$ is an ideal in A. More generally $\varphi^{-1}(I)$ is an ideal in A for any ideal $I \subset B$.*
(2) *Assume $\varphi : A \to B$ is* surjective.
Then the map $J \mapsto \varphi^{-1}(J)$ is a bijection from the set of ideals of B to the set of ideals of A containing $\ker(\varphi)$.

PROOF. (1) The first statement follows from the second by putting $I = \{0\}$.

As J is a subgroup of the additive group of B, also $\varphi^{-1}(J)$ is a subgroup of the additive group of A. For each $a \in \varphi^{-1}(J)$ and each $x \in A$ we have $\varphi(a) \in J$, $\varphi(xa) = \varphi(x)\varphi(a) \in J$ and $\varphi(ax) = \varphi(a)\varphi(x) \in J$. Thus $ax, xa \in \varphi^{-1}(J)$.

(2) (Injectivity) Let J be an ideal in B. Then $\varphi^{-1}(J)$ is an ideal of A, and obviously $\varphi^{-1}(J)$ contains $\varphi^{-1}(0) = \ker(\varphi)$. Furthermore by definition $\varphi(\varphi^{-1}(J)) = J$. Thus $J \mapsto \varphi^{-1}(J)$ is injective.

(Surjectivity) Let I be an ideal of A containing $\ker(\varphi)$. To show that $J \mapsto \varphi^{-1}(J)$ is surjective, we need to find an ideal $J \in B$ with $\varphi^{-1}(J) = I$. Put $J := \varphi(I) \subset B$. First we have to show that J is an ideal in B. As φ is a homomorphism of additive groups, $J = \varphi(I)$ is a subgroup of the additive group of B. Now let $y \in J$ and $b \in B$. Then we can write $y = \varphi(x)$ with $x \in I$ and $b = \varphi(a)$ with $a \in A$. Then, as I is an ideal, we have $xa \in I$ and thus $\varphi(x)\varphi(a) = \varphi(xa) \in \varphi(I) = J$ and similarly $\varphi(a)\varphi(x) \in J$. Thus J is an ideal in B.

We claim that $\varphi^{-1}(J) = I$. Obviously $\varphi^{-1}(J) = \varphi^{-1}(\varphi(I)) \supset I$. We need to see the other inclusion. Thus let $z \in \varphi^{-1}(J)$. Then $\varphi(z) \in J = \varphi(I)$. Thus there exists $y \in I$ with $\varphi(z) = \varphi(y)$, thus $\varphi(z - y) = 0$. Thus $z - y \in ker(\varphi) \subset I$. Thus $z = y + (z - y) \in I$ as sum of two elements of I. □

REMARK 2.10.
(1) Let I be an ideal in a ring R with 1. If I contains a unit, then $I = R$.
(2) The only ideals of a field K are $\{0\}$ and K.
(3) Let K be a field and $\varphi : K \to R$ be a ring homomorphism, then $\varphi = 0$ or φ is injective.

PROOF. This is all very easy. (1) Let $a \in I \cap R^*$. Then there is an element $b \in R$ with $ab = 1$. Thus $1 \in I$ and for all $x \in R$ we have $x = x1 \in I$. (2) follows from (1) because $K^* = K \setminus \{0\}$. (3) follows from (2) because $ker(\varphi)$ is an ideal of K. □

We said before that ideals in rings are analogous to normal subgroups in groups. If $N \subset G$ is a normal subgroup in a group, then one can form the quotient group G/N. Thus for an ideal I in a ring R we want to form the quotient ring R/I.

Let R be a ring and $I \subset R$ an ideal. As the additive group of R is commutative, I is a normal subgroup in $(R, +)$. Thus we can form the quotient group

$$R/I := \big\{x + I \;\big|\; x \in R\big\}.$$

Here $x + I = \big\{x + a \;\big|\; a \in I\big\}$. As R is commutative, R/I is a commutative group with addition

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

and neutral element $0 + I$.

We will look at this in a slightly different way. On R we have an equivalence relation

$$x \sim y \iff x - y \in I.$$

(Check this is an equivalence relation). The equivalence classes are precisely the $x+I$, $x \in R$. Thus R/I is the set of equivalence classes for this equivalence relation. We will usually write $[x]$ instead of $x + I$ for the equivalence class. We have a canonical surjective group homomorphism $\pi : R \to R/I; x \mapsto [x]$ for the additive group. Now we want to see that R/I is also in a natural way a ring and π is a surjective ring homomorphism.

THEOREM 2.11. *Let R be a ring and let $I \subset R$ be an ideal. R/I with the operations*

$$[x] + [y] = [x + y], \quad [x][y] = [xy]$$

is a ring and the natural projection $\pi : R \to R/I; x \mapsto [x]$ is a surjective ring homomorphism with kernel I.

PROOF. We already know that $(R/I, +)$ is a commutative group, and π a group homomorphism. We need to show that the product is well defined (i.e. does not depend on the choice of x and y). If $[x] = [x']$ and $[y] = [y']$, then $x - x', y - y' \in I$. Thus

$$xy - x'y' = (x - x')y - x'(y' - y) \in I,$$

thus $[x'y'] = [xy]$ and the product is well-defined. The associativity and the distributive law for R/I follow directly from that for R. I check this for the associativity:

$$[x]([y][z]) = [x][yz] = [xyz] = [xy][z] = ([x][y])[z].$$

We already know that π is a group homomorphism for the additive groups and we see $\pi(xy) = [xy] = [x][y] = \pi(x)\pi(y)$. □

COROLLARY 2.12. *There is a bijection*

$$\big\{\textit{ideals of } R/I\big\} \to \big\{\textit{ideals of } R \textit{ containing } I\big\}.$$

So we have seen that we can quotient out a ring by an ideal. As $\pi : R \to R/I$ is a ring homomorphism, this shows that the ideals of R are precisely the kernels of ring homomorphisms starting from R. So ideals have two important functions: They are the kernels of ring homomorphisms, and we can take the quotient ring by ideals.

Now we want to show that if we have a homomorphism $\varphi : A \to B$ of rings, and I an ideal contained in the kernel of φ, then we get an induced ring homomorphism $\overline{\varphi} : A/I \to B$. We say that φ can be *factored through* A/I. This property in fact determines A/I up to isomorphism.

THEOREM 2.13. *(Universal property). Let $\varphi : A \to B$ be a ring homomorphism and let $I \subset A$ be an ideal contained in $ker(\varphi)$. Then there exists a unique ring homomorphism $\overline{\varphi} : A/I \to B$, so that the diagram*

$$\begin{array}{ccc} A & \xrightarrow{\varphi} & B \\ \downarrow{\scriptstyle \pi} & \nearrow{\scriptstyle \overline{\varphi}} & \\ A/I & & \end{array}$$

commutes. Furthermore $ker(\overline{\varphi}) = ker(\varphi)/I$ and $\overline{\varphi}(R/I) = \varphi(R)$.

PROOF. Again this is very easy. $\varphi = \overline{\varphi} \circ \pi$ means that $\overline{\varphi}([x]) = \varphi(x)$ for all $x \in R$. Thus $\overline{\varphi}$ is unique. Now we need to show that this formula gives a well-defined ring homomorphism, which is again a straightforward verification. If $[x] = [y]$, then $x - y \in I$. Thus $\varphi(x - y) = 0$ and $\varphi(x) = \varphi(y)$. This shows $\overline{\varphi}$ is well-defined.

$$\overline{\varphi}([x] + [y]) = \overline{\varphi}([x + y]) = \varphi(x + y) = \varphi(x) + \varphi(y) = \overline{\varphi}([x]) + \overline{\varphi}([y]),$$

and similarly $\overline{\varphi}([x][y]) = \overline{\varphi}([x])\overline{\varphi}([y])$. Thus $\overline{\varphi}$ is a ring homomorphism. $\overline{\varphi}(R/I) = \varphi(R)$ is obvious from the definition and

$$[x] \in ker(\overline{\varphi}) \iff x \in ker(\varphi) \iff [x] \in ker(\varphi)/I.$$

□

The most important case of this is the Homomorphism Theorem: for a surjective ring homomorphism $\varphi : A \to B$, B is isomorphic to A/I.

COROLLARY 2.14. *(Homomorphism Theorem) Let $\varphi : A \to B$ be a surjective ring homomorphism with kernel I. Then the map*

$$\overline{\varphi} : A/I \to B; [x] \mapsto \varphi(x)$$

is a ring isomorphism.

PROOF. By the universal property of the quotient $\overline{\varphi}$ is a surjective ring homomorphism, with kernel $I/I = [0]$. Thus it is an isomorphism. □

The Homomorphism Theorem is very useful, we will apply it very often.

Exercises.

(1) Let R be a ring and let I and J be ideals in R. Which of the following are ideals of R

$$I + J := \{u + v \mid u \in I,\, v \in J\},$$
$$I \cap J,$$
$$I \cup J?$$

Give a proof or a counterexample.

(2) Let R be a noncommutative ring and $a \in R$. Show by example that $aR := \{ar \mid r \in R\}$ is not always an ideal.

(3) Let R be a ring, $I \subset R$ an ideal.
   (a) Show that $r(I) := \{r \in R \mid ra = 0 \text{ for all } a \in I\}$ is an ideal in R.
   (b) Show that $[R : I] := \{x \in R \mid xr \in I \text{ for all } r \in R\}$ is an ideal in R.

(4) Prove that a subring of an integral domain is an integral domain.

(5) Is there an integral domain with precisely 10 elements?

(6) Let R be a ring such that $x^3 = x$ for every $x \in R$. Show that R is a commutative ring.

(7) Let R be a ring with 1 and let $\varphi : R \to S$ be a surjective homomorphism of rings. Show that $\varphi(1)$ is the unit element of S.

(8) Let k be a field. Find all automorphisms φ of $k[x]$ with the property that $\varphi(a) = a$ for all $a \in k$.

(9) Prove that every ideal in $\mathbb{Z}[i]$ contains a nonzero integer.

(10) Show that the kernel of the homomorphism $\varphi : \mathbb{C}[x, y] \to \mathbb{C}[t] : x \mapsto t^2, y \mapsto t^3$ is the principal ideal generated by $y^2 - x^3$.

(11) Let R be a ring and I an ideal in R. Let $\sqrt{I} := \{x \in R \mid x^n \in I \text{ for some } n > 0.\}$. Show
   (a) $\sqrt{I}$ is an ideal in R which contains I. $\sqrt{I}$ is usually called the radical of I.
   (b) $\sqrt{\sqrt{I}} = \sqrt{I}$.

(12) Let R be a ring and let A, B be ideals in R such that $A \cap B = \{0\}$. Show that for all $a \in A$, $b \in B$, $ab = 0$.

(13) Let R be a ring (commutative with 1) and let S be a subring. The *conductor* $C(S)$ of S in R is the set of all $\alpha \in R$ such that $\alpha R \subset S$.
   (a) Prove that $C(S)$ is an ideal of R and also an ideal of S.
   (b) Prove that $C(S)$ is the largest ideal of S which is also an ideal of R.
   (c) Determine $C(S)$ in the following cases:
      (i) $R = \mathbb{C}[x]$, $S = \mathbb{C}[x^2, x^3]$.
      (ii) $R = \mathbb{Z}[\omega]$. $\omega = \frac{-1+\sqrt{-3}}{2}$, $S = Z[\sqrt{-3}]$.

(14) Let I, J be ideals of a ring, such that $I + J = R$. Prove that $IJ = I \cap J$.

3. Prime Ideals and Maximal Ideals

From now on until the end of the lectures a ring is a commutative ring with 1 and a homomorphism $\varphi : A \to B$ of rings satisfies $\varphi(1) = 1$.

We now want to look at the two most important classes of ideals. We have learned in the last section, that for an ideal I in a ring R a natural thing to do is to form the quotient ring R/I. The nicest ideals will be those for which the quotient is the nicest. The nicest rings are the fields, where we can compute like in $\mathbb{Q}$, followed by the integral domains, where we can compute like in $\mathbb{Z}$. A prime ideal in a ring R will be an ideal P so that R/P is an integral domain, and a maximal ideal M in R will be such that R/M is a field.

DEFINITION 3.1. Let R be a ring (remember this means commutative with 1). An ideal $P \subsetneq R$ is called a *prime ideal* if for all $a, b \in R$ with $ab \in P$, we have $a \in P$ or $b \in P$.

EXAMPLE 3.2. Let $m \in \mathbb{Z}_{>0}$. Then $\langle m \rangle = m\mathbb{Z}$ is a prime ideal in $\mathbb{Z}$ if and only if m is a prime number.

A maximal ideal is a proper ideal that does not fit into any bigger proper ideal.

DEFINITION 3.3. Let R be a ring. An ideal $M \subsetneq R$ is called *maximal* if there is no ideal I with $M \subsetneq I \subsetneq R$.

THEOREM 3.4. (1) *Let R be a ring. An ideal $P \subset R$ is a prime ideal if and only if R/P is an integral domain.*
(2) *Let R be a ring. An ideal $M \subset R$ is a maximal ideal if and only if R/M is a field.*

PROOF. (1) As R is commutative with 1 also R/P is commutative with 1. For $a, b \in R$, $[a] = 0$ if and only if $a \in P$; and $0 = [a][b] = [ab]$ only if $ab \in P$. Thus $[a]$ is a zero divisor if and only if $a \notin P$ and there is an element $b \notin P$ with $ab \in P$.

(2) Suppose first M is an ideal in R such that R/M is a field. As R/M is a field, the only ideals in R/M are 0 and R/M. By the bijection of ideals in R/M and ideals in R containing M, the only ideals in R containing M are M and R. Thus M is a maximal ideal. If on the other hand M is maximal ideal, by the same bijection, the only ideals in R/M are 0 and R/M. Thus the theorem follows from the next lemma. □

LEMMA 3.5. *A commutative ring R with 1 whose only ideals are $\{0\}$ and R is a field.*

PROOF. For every $a \in R \setminus \{0\}$ we need to find an element $b \in R$ with $ab = 1$. $\langle a \rangle := \{ab \mid b \in R\}$ is an ideal in R, thus $\langle a \rangle = 0$ or $\langle a \rangle = R$. Since $0 \neq a = 1a \in \langle a \rangle$ we see that $\langle a \rangle = R$. Therefore $1 \in \langle a \rangle$, i.e. there exists $b \in R$ with $ab = 1$. □

As a field is in particular an integral domain we also get.

COROLLARY 3.6. *Every maximal ideal is a prime ideal.*

EXAMPLE 3.7. (1) If K is a field, then $\{0\}$ is a maximal ideal in K.
(2) We have seen that $\langle m\rangle$ is a prime ideal in $\mathbb{Z}$ if and only if m is a prime number. But in this case $\langle m\rangle$ is even a maximal ideal because $\mathbb{Z}/\langle m\rangle$ is a field.

Exercises.

(1) Let R be a commutative ring with 1. An element $r \in R$ is called nilpotent if there exists an $n \in \mathbb{Z}_{>0}$ with $r^n = 0$. Let
$$N(R) := \big\{r \in R \;\big|\; r \text{ is nilpotent}\big\}.$$
   (a) Determine $N(\mathbb{Z}/4\mathbb{Z})$.
   (b) Show that $N(R)$ is an ideal in R.
   (c) Show that $N(R) \subset P$ for every prime ideal $P \subset R$.
   (d) Show that 0 is the only nilpotent element of $R/N(R)$.
(2) Let k be a field and $f \in k[x]$. Let $R := k[x]/\langle f\rangle$. Show that $N(R) = 0$ (see previous exercise) if and only if f is not divisible by the square of a polynomial.
(3) Determine the maximal ideals of $\mathbb{R}[x]/(x^2)$ and $\mathbb{R}[x]/(x^2 + x + 1)$.
(4) Prove that the ideal $\langle x + y^2, y + x^2 + 2xy^2 + y^4\rangle$ in $\mathbb{C}[x, y]$ is a maximal ideal.
(5) Let R be a ring, and let I be an ideal of R. Let $\overline{R} := R/I$. For an ideal $M \subset R$ containing I let $\overline{M} = M/I$. Show that $\overline{M}$ is maximal in $\overline{R}$ if and only if M is maximal in R.
(6) Give an example of a ring in which some prime ideal is not a maximal ideal.
(7) Let R be a commutative ring with 1 and let I be an ideal that is contained in a finite union $p_1 \cup \ldots \cup p_n$ of prime ideals of R. Show that $I \subset p_i$ for some i.
(8) Let $f : R \to S$ be a surjective ring homomorphism with kernel K. Show:
   (a) If P is a prime ideal of R containing K, then $f(P)$ is a prime ideal of S.
   (b) If Q is a prime ideal of S, then $f^{-1}(Q)$ is a prime ideal of R containing K.
(9) Let R be a ring with 1, M an ideal in R. Suppose every element of $R \setminus M$ is a unit in R. Show that M is the unique maximal ideal of R.
(10) Show: The ring $2\mathbb{Z}$ of even integers contains a maximal ideal M, such that $2\mathbb{Z}/M$ is not a field.

4. Polynomial rings over a field

We have already introduced the polynomial rings $R[x]$ in the first section. In this section we want to study the case of polynomial rings over a field in more detail. In this section let k be a field. We want to start out by looking at questions of divisibility in $k[x]$. So we first define this in general.

DEFINITION 4.1. Let R be an integral domain. Let $a, b \in R$. We say that a *divides* b, denoted $a|b$, if there exists an element $c \in R$ with $b = ac$. Otherwise we write $a \not| b$.

The relation "divides" has a number of obvious properties.

REMARK 4.2. $a|b$ is equivalent to $b \in \langle a \rangle$, in particular $a|b$ and $b|c$ implies $a|c$; and $a|b$ and $a|c$ implies $a|(b \pm c)$.

Questions of divisibility of polynomials over k are governed by the algorithm of division with rest, that works in $k[x]$ in a very similar way as for integers.

THEOREM 4.3. *(Division with rest) Let $f, g \in k[x]$ with $g \neq 0$. There are unique $q \in k[x]$ and $r \in k[x]$ such that*

$$f = qg + r, \qquad \text{with } deg(r) < deg(g) \text{ or } r = 0.$$

PROOF. Existence: If $f = 0$ or $deg(f) < deg(g)$, we can put $q = 0$ and $r = f$. Thus assume that $m := deg(f) \geq deg(g)$ and make induction on m.

Let a be the leading coefficient of f and b the leading coefficient of g. Then the coefficient of x^m in

$$\overline{f} := f - (\frac{a}{b} x^{m-deg(g)})g$$

is $a - ab/b = 0$, thus $\overline{f}$ has degree $\leq m-1$, and by induction we can write $\overline{f} = q'g + r'$ with $r' = 0$ or $deg(r') < deg(g)$. But then we put $r := r'$ and $q = q' + (\frac{a}{b} x^{m-deg(g)})$, and the claim is satisfied.

Uniqueness: If $qg + r = q'g + r'$ with the above properties, then $(q - q')g = r - r'$. $q \neq q'$ would imply

$$deg(r - r') = deg(q - q') + deg(g) \geq deg(g),$$

a contradiction to $deg(r), deg(r') < deg(g)$. Thus $q = q'$ and thus $r = r'$. □

EXAMPLE 4.4. Note that the proof of the Theorem actually gives us an algorithm for division with rest.

$$\begin{array}{rll} (x^3 + 4x^2 + x + 1) = (x+3)(x^2 + x - 5) + 3x + 16 & & \\ x^3 + x^2 - 5x & x & \\ 3x^2 + 6x + 1 & & \\ 3x^2 + 3x - 15 & 3 & \\ 3x + 16 & & \end{array}$$

In $\mathbb{Z}$ we can talk about the greatest common divisors. Now we introduce greatest common divisors in any integral domain R.

DEFINITION 4.5. Let R be an integral domain and let $a_1, \dots, a_n \in R$. An element $r \in R \backslash \{0\}$ is called a *common divisor* of $a_1, \dots, a_n$, if $r|a_i$ for $i = 1, \dots, n$. An element $r \in R$ is called a *greatest common divisor* of $a_1, \dots, a_n$, if r is a common divisor, and $s|r$ for any other common divisor. $a_1, \dots, a_r$ are called *relatively prime*, if 1 is a greatest common divisor.

By definition the greatest common divisor of $a_1, \dots, a_r$ is not unique, but it is unique up to multiplication by a unit. In $k[x]$ the division algorithm gives an algorithm to find the greatest common divisor.

REMARK 4.6. Let $f, g \in k[x]$. We put $r_0 := f$, $r_1 := g$. Then by the division algorithm we find $f = q_1 g + r_2$, $g = q_2 r_2 + r_3$ and inductively

$$r_{i-1} = q_i r_i + r_{i+1}, \quad \text{with } q_i \in k[x],\ r_{i+1} \in k[x], \quad \text{and } deg(r_{i+1}) < deg(r_i).$$

The process stops when $r_{n-1} = q_n r_n$ and $r_n \neq 0$.

We claim that then r_n is a greatest common divisor of f and g: The equation $r_{i-1} = q_i r_i + r_{i+1}$ implies sucessively that $r_n|r_{n-1},\ r_n|r_{n-2}, \dots, r_n|g,\ r_n|f$. Thus r_n is a common divisor of f and g. On the other hand, if t is a common divisor of f and g, then the same equation implies sucessively that $t|r_2,\ t|r_3, \dots, t|r_n$. Thus r_n is a greatest common divisor. This algorithm is called the *Euclidean algorithm.*

It implies one further result, which we will use in the future (in the proof of the Theorem of the Primitive Element). If K is another field, so that $k \subset K$ is a subring then we call k a subfield of K. It is straightforward to see that then $k[x]$ is a subring of $K[x]$.

COROLLARY 4.7. *Let K be a field and $k \subset K$ a subfield. Let $f, g \in k[x]$. Let h be a greatest common divisor of f and g in $K[x]$ and assume that its leading coefficient is in k. Then $h \in k[x]$.*

PROOF. Let l be the greatest common divisor of f and g computed via the Euclidean algorithm. Then $l \in k[x]$ because in the division with rest of two elements of $k[x]$ both quotient and rest are in $k[x]$ and the Euclidean algorithm is just repeatedly applying division with rest. Note the the greatest common divisor in $K[x]$ is well-defined up to multiplication by a constant $a \in K$. Let $h \in K[x]$ be a greatest common divisor of f, g in $K[x]$. Let h_n be the leading coefficient of h, and assume that $h_n \in k$. We know $h = al$ for some $a \in K \setminus \{0\}$. Let l_n be the leading coefficient of l. Then $l_n \in k$. Thus $a = h_n/l_n \in k$. Thus $h = al \in k[x]$. □

A very important property of polynomials is that we can evaluate them at any element of k and more generally at any element of a field K that contains k as a subfield, i.e. we can substitute for x any element of k or of K.

DEFINITION 4.8. let $f = \sum_{i=0}^{n} a_i x^i \in k[x]$ and let R be a ring that contains k as a subring. The *value* of f in $s \in R$ is

$$f(s) := \sum_{i=0}^{n} a_i s^i \in R.$$

It is straightforward to check that

$$(f+g)(s) = f(s) + g(s), \quad fg(s) = f(s)g(s).$$

Thus we get a ring homomorphism

$$ev_s : k[x] \to R, f \mapsto f(s).$$

ev_s is called the *evaluation homomorphism* at s. An element $s \in R$ is called a *zero* of f if $f(s) = 0$.

REMARK 4.9. For the moment the most important case will be that $R = k$, so that we can evaluate a polynomial in $k[x]$ at any element of k. But we shall see later that the general case is important.

Now we want to use the division with rest to study zeros of polynomials in $k[x]$. If $a \in k$ is a zero of $f \in k[x]$, then we can divide f by $x - a$. This is then used to show that if f has degree n, then it can have at most n zeros in k.

THEOREM 4.10. *Let $f \in k[x]$ and $a \in k$. Then a is a zero of f, if and only if $(x-a)|f$, i.e. if $f = (x-a)g$ for some $g \in k[x]$.*

PROOF. If $(x-a)$ divides f, then $f = (x-a)g$, thus $f(a) = (a-a)g(a) = 0$. So a is a zero of f.

Thus assume $f(a) = 0$. If $f = 0$, we can put $g = 0$. Otherwise by division with rest we can write $f = (x-a)g + r$ with $deg(r) = 0$ or $r = 0$. Thus $r \in k$. Evaluating at a we get

$$0 = f(a) = (a-a)g(a) + r = r,$$

i.e. $f = (x-a)g$. □

THEOREM 4.11. *Let $f \in k[x] \setminus \{0\}$. Then f has at most $deg(f)$ zeros in k.*

PROOF. We show this by induction on the degree of f. If f has degree 0, then f is constant, and thus has no zero. Now let f have degree $n+1$. If f has no zero in k, then the claim is trivially true. Otherwise let $a \in k$ be a zero of f. Then we can write $f = (x-a)g$ with $deg(g) = n$. Thus by induction g has at most n zeros in k. If $b \neq a$ is a zero of f, then $0 = (b-a)g(b)$, thus $g(b) = 0$. So f has at most $n+1$ zeros in k. □

Exercises.

(1) Find the greatest common divisors of the following polynomials in $\mathbb{Q}[x]$.
   (a) $x^3 - 6x^2 + x + 4$ and $x^5 - 6x + 1$.

(b) x^2+1 and x^6+x^3+x+1.

(2) Let k be a field and let $a_0, \dots, a_n \in k$ be distinct elements and let , $b_0, \dots, b_n \in k$. Show: There is at most one polynomial $f \in k[x]$ of degree n with $f(a_i) = b_i$ for $i = 0, \dots, n$.

(3) Let $a_0, \dots, a_{n-1} \in \mathbb{Z}$ and let

$$f = x^n + \sum_{i=0}^{n-1} a_i x^i \in \mathbb{Q}[x]$$

(note that the leading coefficient is 1). Let $b \in \mathbb{Q}$ be a zero of f. Show that $b \in \mathbb{Z}$.

(4) Let $f, g \in k[x]$ be polynomials with coefficients in a field k. Assume the both f and g factor into linear factors i.e. $f = (x+a_1) \cdot \ldots \cdot (x+a_n)$, $g = (x+b_1) \cdot \ldots \cdot (x+b_m)$. Then the greatest common divisor is the product of the common linear factors.

5. Euclidean rings and principal ideal domains

An important property of the integers is that in the integers we can do division with rest, i.e. if a, b are integers, we can write

$$a = tb + r, \quad t, r \in \mathbb{Z}, \quad 0 \leq r < b.$$

r is the rest of the division of a by b. In the last section we have seen that also in polynomial rings over a field we have division with rest. Now we want to formalize this. The rest r should in some sense be smaller then b, in general rings we have to measure this by a function $d : R \setminus \{0\} \to \mathbb{Z}_{>0}$.

Definition 5.1. An integral domain R is called a *Euclidean ring* if there is a function $d : R \setminus \{0\} \to \mathbb{Z}_{\geq 0}$ with the following properties.

(1) For all a, b in $R \setminus \{0\}$ there exist $q, r \in R$ such that $a = qb + r$ with either $r = 0$ or $d(r) < d(b)$.

(2) If $b \in R \setminus \{0\}$ is not a unit, then $d(ab) > d(a)$ for all $a \in R \setminus \{0\}$.

Example 5.2. (1) $\mathbb{Z}$ with $d(n) = |n|$ is a Euclidean ring.

(2) Let again $\mathbb{Z}[i] := \{n + mi \mid n, m \in \mathbb{Z}\} \subset \mathbb{C}$ be the ring of Gaussian integers. We claim that $\mathbb{Z}[i]$ is a Euclidean ring with $d(n+im) = n^2 + m^2$. We can extend d to $\mathbb{C}$, by $d(a+ib) = a^2 + b^2$, d is just the square of the complex absolute value. In particular $d(a+ib) \neq 0$ for $a + ib \neq 0$. We then have $d(zw) = d(z)d(w)$ for all $z, w \in \mathbb{C}$. If $z, w \in \mathbb{Z}[i] \setminus \{0\}$, let $z/w = a + ib$ be the quotient in $\mathbb{C}$. Choose $n, m \in \mathbb{Z}$ such that $|a - m| \leq 1/2$, $|b - n| \leq 1/2$. Then

$$d(z/w - (m + in)) = (a-m)^2 + (b-n)^2 \leq 1/2.$$

Thus we can put $r := z - (m + in)w$ as the rest of the division, and get $d(r) = d(w)d(z/w - (m+in)) < d(w)$.

(3) The most important example of a Euclidean ring is that of a polynomial ring over a field. Let k be a field, and for $f \in k[x] \setminus \{0\}$ put $d(f) = deg(f)$. Then $(k[x], d)$ is a Euclidean ring. This is just a reformulation of Theorem **??**.

Euclidean rings have another nice property. The set of their ideals is very simple: every ideal $I \subset R$ is a principal ideal.

DEFINITION 5.3. Let R be an integral domain. Recall that an ideal of the form $\langle a \rangle = \{ar \mid r \in R\}$ for an $a \in R$ is called a *principal ideal*. An integral domain R is called a *principal ideal domain* (PID), if every ideal $I \subset R$ is a principal ideal.

THEOREM 5.4. *A Euclidean ring is a principal ideal domain.*

PROOF. Let (R, d) be a Euclidean ring. We have to show that every ideal $I \subset R$ is a principal ideal. The zero ideal $\{0\} = \langle 0 \rangle$ is principal. Thus let $I \neq \{0\}$ be an ideal. Let $a \in I \setminus \{0\}$ be such that $d(a)$ is the smallest $d(b)$ for $b \in I \setminus \{0\}$. We want to show $I = \langle a \rangle$: Otherwise there is an element $b \in I \setminus \langle a \rangle$. Then division with rest gives

$$b = ta + r, \quad t, r \in R \setminus \{0\}, \quad d(r) < d(a).$$

This is a contradiction to the choice of a, because $r = b - ta \in I$. □

COROLLARY 5.5. *$\mathbb{Z}$, $\mathbb{Z}[i]$, and $k[x]$ for a field k are principal ideal domains.*

We see in particular that the ideals in $\mathbb{Z}$ are precisely the $m\mathbb{Z}$ that we already know and the quotient rings are precisely the $\mathbb{Z}/m\mathbb{Z}$.

In general integral domains greatest common divisors do not always need to exist. However in a PID they do.

THEOREM 5.6. *Let R be a principal ideal domain. Let $a_1, \ldots, a_r \in R$. Then there is greatest common divisor d of $a_1, \ldots, a_r$ of the form $d = a_1x_1 + \ldots + a_rx_r$ with $x_i \in R$.*

PROOF. Choose $d \in R$ such that $\langle d \rangle = \langle a_1, \ldots, a_r \rangle$. Then by definition $d = a_1x_1 + \ldots + a_rx_r$ with $x_i \in R$. We claim that d is a greatest common divisor. By definition $a_i \in \langle d \rangle$ for all i, thus $d | a_i$. On the other hand if $e \in R$ divides all a_i, then $e | a_1x_1 + \ldots + a_rx_r = d$. □

EXAMPLE 5.7. Let $a, b \in \mathbb{Z}$, then $\langle a, b \rangle = \langle gcd(a, b) \rangle$, e.g. $\langle 4, 6 \rangle = \langle 2 \rangle$ and $\langle 4, 7 \rangle = \mathbb{Z}$.

Exercises.

(1) Prove or disprove the following statement: The ring $\mathbb{Z}[x]$ is a principal ideal domain.
(2) If $a + bi \in \mathbb{Z}[i]$ is not a unit, show that $a^2 + b^2 > 1$.
(3) Show that $\mathbb{Z}[\sqrt{2}]$ is an Euclidean domain.
(4) Give an example that division with rest need not be unique in a Euclidean domain.

(5) Let $m, n \in \mathbb{Z}$. Show that their greatest common divisor in $\mathbb{Z}$ is the same as their greatest common divisor in $\mathbb{Z}[i]$.

6. Irreducibility of polynomials

An element in an integral domain will be called irreducible, if whenever we can write it as a product, one of the factors has to be a unit. In $\mathbb{Z}$ the prime numbers are the irreducible elements. In this section with will study questions of irredicibility of polynomials in $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$.

We start by introducing irreducible elements in an integral domain R, which are generalizations of the prime numbers in $\mathbb{Z}$. An irreducible element is one that cannot in a nontrivial way be written as a product of other elements

Definition 6.1. An element $q \in R \setminus \{0\}$ is called *irreducible* if q is not a unit and for any $a, b \in R$ with $q = ab$ either a or b is a unit. An element of R is called reducible, if it is not irreducible.

Remark 6.2. A prime number $p \in \mathbb{Z}_{>0}$ is irreducible in $\mathbb{Z}$. Prime numbers $p \in \mathbb{Z}_{>0}$ have another important property: Let $a, b \in \mathbb{Z}$. If $p|ab$, then $p|a$ or $p|b$. Elements in an integral domain with this property are also called prime elements. In a general integral domain prime elements and irreducible elements are not the same.

Example 6.3.
(1) A number $q \in \mathbb{Z}$ is irreducible if $q = \pm p$ for p a prime number.
(2) A field has no irreducible elements.
(3) If k is a field and $a \in k \setminus \{0\}$, $b \in k$, then $ax + b$ is an irreducible element of $k[x]$: If $ax+b = fg$, then $deg(f)+deg(g) = 1$, thus $deg(f) = 0$ or $deg(g) = 0$, thus $f \in k^*$ or $g \in k^*$.
(4) The irreducible elements in $R[x]$ are called irreducible polynomials.

Now we want to see that in a principal ideal domain irreducible elements generate maximal ideals.

Proposition 6.4. *Let R be a principal ideal domain and $p \in R$ an irreducible element. Then $\langle p \rangle$ is a maximal ideal, and $R/\langle p \rangle$ is a field*

Proof. Let $p \in R$ be an irreducible element. We know that every ideal in R is of the form $\langle a \rangle$. Thus let $\langle p \rangle \subset \langle a \rangle$. Then $p \in \langle a \rangle$, thus $p = ab$ for some $b \in R$. Then, as p is irreducible, either a is a unit and $\langle a \rangle = R$ or b is a unit and $\langle a \rangle = \langle p \rangle$. Thus $\langle p \rangle$ is maximal. □

This gives us a way to construct new fields. We will use this in the chapter on fields.

Corollary 6.5. *Let k be a field and let f in $k[x]$ be an irreducible polynomial. Then $k[x]/\langle f \rangle$ is a field containing k as a subfield (if we identify with the image of the constant polynomials in $k[x]/\langle f \rangle$).*

We now want to study the irreducibility of polynomials over $\mathbb{Z}$ and $\mathbb{Q}$.

DEFINITION 6.6. A polynomial $f = \sum_{i=0}^{n} a_i x^i \in \mathbb{Z}[x]$ is called *primitive* if the coefficients $a_0, \dots, a_n$ are relatively prime.

LEMMA 6.7. *(Lemma of Gauss) Let $f, g \in \mathbb{Z}[x]$ be primitive polynomials. Then also fg is primitive.*

PROOF. Let $f = \sum_{i=0}^{n} a_i x^i$, $g = \sum_{i=0}^{m} b_i x^i$. Suppose that fg is not primitive. Then there is a prime number p which is a common divisor of the coefficients of fg. Let a_i be the lowest coefficient of f that p does not divide and b_j the lowest coefficient of g that p does not divide. Then in fg the coefficient of x^{i+j} is

$$c_{i+j} = a_i b_j + \sum_{k>i} a_k b_{i+j-k} + \sum_{k<i} a_k b_{i+j-k}.$$

Then p divides b_{i+j-k} for $k > i$ and p divides a_k for $k < i$. Thus p divides both sums. As p does not divide $a_i b_j$, we get p does not divide c_{i+j}. □

The following theorem is also sometimes called the Gauss Lemma.

THEOREM 6.8. *(Gauss Lemma) Let $f \in \mathbb{Z}[x]$ be a non-constant primitive polynomial. Then f is irreducible in $\mathbb{Z}[x]$ if and only if it is irreducible in $\mathbb{Q}[x]$.*

PROOF. "$\Longleftarrow$" If f is reducible in $\mathbb{Z}[x]$, then we have $f = gh$, with g, h both not units in $\mathbb{Z}[x]$. If $deg(g) = 0$, then g is a nonunit in $\mathbb{Z}$, which is a common factor of all the coefficients of fg, contradicting the fact that f is primitive. The same happens if $deg(h) = 0$. Thus $f = gh$ with $deg(g) > 0, deg(h) > 0$. Thus it is also reducible in $\mathbb{Q}[x]$.

"$\Longrightarrow$" Suppose $f = gh$ where $g, h \in \mathbb{Q}[x]$ are polynomials of positive degree. By clearing denominators and dividing by the greatest common divisor of the coefficients, we can write $f = \frac{a}{b} g'h'$, where g', h' are primitive polynomials in $\mathbb{Z}[x]$ and $a, b \in \mathbb{Z}$ are relatively prime. Thus $bf = ag'h'$. But both f and $g'h'$ are primitive. So on the left hand side the greatest common divisor of the coefficients is b and on the right hand side it is a. Thus $a = \pm b$, and thus $f = \pm g'h'$ is reducible in $\mathbb{Z}[x]$. □

Now we give a useful criterion for the irreducibility of polynomials.

THEOREM 6.9. *(Eisenstein's criterion) Let $f = \sum_{i=0}^{n} a_i x^i$ be a primitive polynomial in $\mathbb{Z}[x]$ of positive degree. Assume there is a prime number p with $p|a_0, p|a_1, \dots, p|a_{n-1}$ but $p \not| a_n$ and $p^2 \not| a_0$.*

Then f is irreducible over $\mathbb{Z}$, and over $\mathbb{Q}$.

PROOF. Assume $f = gh$ with $g, h \in \mathbb{Z}[x]$. We have to show $g = \pm 1$ or $h = \pm 1$. Let $g = \sum_{i=0}^{k} b_i x^i$ with $b_k \neq 0$ and $h = \sum_{i=0}^{l} c_i x^i$ with $c_l \neq 0$. As $a_0 = b_0 c_0$ and $p|a_0$ and $p^2 \not| a_0$, we see that p divides exactly one of b_0, c_0. Assume $p|b_0$ and $p \not| c_0$. As

$a_n = b_k c_l$ and $p \not| a_n$ we see that $p \not| b_k$. Thus there exists a maximal $j \in \{1, \dots, k\}$ such that $p|b_i$ for $i < j$ and $p \not| b_j$. Putting $c_i := 0$ for $i > l$, we get

$$a_j = b_j c_0 + b_{j-1} c_1 + \ldots + b_0 c_j.$$

By definition p does not divide $b_j c_0$, but it divides each other summand. Therefore $p \not| a_j$. Thus $j = n$ and thus $k = n$. Thus $deg(g) = n$ and therefore h is a constant polynomial. As f is primitive, we get $h = \pm 1$. □

EXAMPLE 6.10. (1) $f := x^5 - 9x + 3$ is irreducible in $\mathbb{Z}[x]$ by Eisenstein's criterion with $p = 3$. Thus by Gauss Lemma it is irreducible in $\mathbb{Q}[x]$.
(2) For p a prime number and $n \in \mathbb{Z}_{>0}$ the polynomial $x^n - p$ is irreducible over $\mathbb{Q}$.

To later study examples of field extensions it is important to be able to check the irreducibility of polynomials in $k[x]$. The most important case is again that of $k = \mathbb{Q}$. We give a few more methods to check irreducibility.

REMARK 6.11. If $f \in k[x]$ is a polynomial of degree at most 2 or 3, then f is irreducible if and only if it has no zero in k (because if f is reducible, then f must have a factor of degree 1). Let in particular $f = x^n + \sum_{i=0}^{n-1} b_i x^i \in \mathbb{Z}[x]$ be a monic polynomial. Then it is shown in an exercise that any zero $a \in \mathbb{Q}$ of f must lie in $\mathbb{Z}$. It is easy to check (exercise) that then $f/(x-a) \in \mathbb{Z}[x]$ and that $a|b_n$. So if $n \leq 3$ we can check the irreducibility very fast by showing that none of the divisors of b_n is a zero of f.

Another possiblity is to substitute something else for the variable of f.

REMARK 6.12. Let $f = \sum_i a_i x^i \in k[x]$, let $a \in k$. Then f is irreducible in $k[x]$ if and only if $f(x+a) := \sum_i a_i (x+a)^i$ is irreducible in $k[x]$): It obvious that $\sigma_a : k[x] \to k[x], g \mapsto g(x+a)$ is an isomorphism of rings with inverse σ_{-a}. Thus f is irreducible if and only if $f(x+a)$ is irreducible.

EXAMPLE 6.13. Let p be a prime number. Then $f = x^{p-1} + x^{p-2} + \ldots + x + 1$ is irreducible in $\mathbb{Q}[x]$.

Note that $(x-1)f = x^p - 1$. Thus $x\sigma_1(f) = (x+1)^p - 1$ and thus $\sigma_1(f) = \sum_{i=1}^{p} \binom{p}{i} x^{i-1}$. It is easy to check that $p|\binom{p}{i}$ for $1 \leq i \leq p-1$ and $p^2 \not| \binom{p}{1} = p$. Thus $\sigma_1(f)$ is irreducible by Eisensteins criterion, and thus f is irreducible over $\mathbb{Q}$.

Exercises.

(1) Prove that
   (a) $x^2 + x + 1$ is irreducible in $F_2[x]$.
   (b) $x^2 + 1$ is irreducible in $F_7[x]$.
(2) Let $a \in \mathbb{Q}$ and assume $(x-a)|f$ for $f \in \mathbb{Z}[x]$ a monic polynomial. Show that $a \in \mathbb{Z}$.

(3) Let R be a commutative ring with no nonzero nilpotent elements (i.e. $a^n = 0$ implies $a = 0$). If $f = \sum_{i=0}^{n} a_i x^i \in R[x]$ is a zero divisor, show that there is an element $b \neq 0$ in R with $ba_0 = ba_1 = \ldots ba_n = 0$.
(4) Let R be a commutative ring with 1. Show that a polynomial $f = \sum_{i=0}^{n} a_i x^i \in R[x]$ is a unit in $R[x]$ if and only if a_0 is a unit in R and $a_1, \ldots, a_n$ are nilpotent.
(5) Let k be a field. Show that $k[x_1, x_2]$ is not a principal ideal domain.
(6) Prove that $1 + x + x^3 + x^4$ is irreducible in $k[x]$ for any field k.
(7) Prove that $x^4 + 2x + 2$ is irreducible in $\mathbb{Q}[x]$.
(8) Let p be a prime number, $c \in \mathbb{Q}$. Show $x^p - c$ is irreducible over $\mathbb{Q}$ if and only if it has no zero in $\mathbb{Q}$.
(9) Let k be a field, $a, b \in k$. Show that $f \in k[x]$ is irreducible if and only if $f(ax + b)$ is irreducible.
(10) Prove that the kernel of the homomorphism $\mathbb{Z}[x] \to \mathbb{R}; x \mapsto 1 + \sqrt{2}$ is a principal ideal and find a generator for this ideal.
(11) Prove that the following polynomials are irreducible in $\mathbb{Q}[x]$.
  (a) $x^2 + 27x + 213$.
  (b) $x^3 + 6x + 12$.
  (c) $x^5 - 3x^4 + 3$.
(12) Factor $x^5 + 5x + 5$ into irreducible factors in $\mathbb{Q}[x]$ and $F_2[x]$.
(13) Suppose that a polynomial $x^4 + bx^2 + c \in \mathbb{Q}[x]$ is the product of two factors of degree 2 in $\mathbb{Q}[x]$. What can one say about the coefficients of these factors?
(14) Factor the following polynomials in $\mathbb{Q}[x]$ (Hint: Use reduction modulo 2).
  (a) $x^3 + 2x^2 + 3x + 1$.
  (b) $x^4 + 2x^3 + 3x^2 + 2x + 1$.
(15) Let $p \in \mathbb{Z}$ be a prime number. Let $f = \sum_{i=0}^{2n+1} a_i x^i \in \mathbb{Z}[x]$ be a polynomial of odd degree. Suppose $p \not| a_{2n+1}$, $p^2 | a_0, \ldots, a_n$, $p | a_{n+1}, \ldots, a_{2n}$ and $p^3 \not| a_0$. Show that f is irreducible in $\mathbb{Q}[x]$.

CHAPTER 3

Fields

In this second half of the lecture we want to study fields. We have already said before that the most important class of rings are the fields. Recall that a field is a commutative ring with 1 in which every nonzero element is a unit. This means that we can compute in a field basically like in the rational numbers $\mathbb{Q}$.

Let k be a field and $f \in k[x]$ a polynomial. We want to study field extensions L/k, i.e. L and k are fields and k is a subring of k. Let $f \in k[x]$ be a polynomial. We can also view it as a polynomial in L. It could happen that f has more zeros in L then in k, e.g.

(1) $k = \mathbb{Q}$, $L = \mathbb{R}$, $f = x^2 - 2$,
(2) $k = \mathbb{R}$, $L = \mathbb{C}$, $f = x^2 + 1$.

Assume f has no zero in k. One leading question is whether we can find an extension L/k, so that f has a zero in L. Can we even find L, so that f splits into linear factors over L?

Thus we will be concerned with algebraic extensions of a field k, i.e. we will study fields $L \supset k$ such that every element of L is the zero of a polynomial $f \in k[x]$. It turns out that to such field extensions one can associate a group, the Galois group. Galois theory is the study of field extensions in terms of their Galois groups. In particular we will show the principal theorem of Galois theory, which relates the subgroups of the Galois group to the intermediate fields $k \subset M \subset L$ of a field extension.

For many centuries mathematicians were interested in finding the zeros of polynomials $f \in \mathbb{Q}[x]$. In particular they tried to express the zeros in terms of radicals $\sqrt[n]{\ }$. Already the Babylonians knew how to do this for polynomials of degree 2. In the 16th century formulas were found for polynomials of degrees 3 and 4, and for almost 300 years people tried to find such formulas for degrees 5 and higher. One of the most striking applications of Galois theory was to show that no such formulas exist in general for $n > 4$.

1. Field extensions, degree theorem

Let R be a ring, let $a \in R$ and let $n \in \mathbb{Z}_{>0}$. Recall that we write $n \cdot a = \underbrace{a + \ldots + a}_{n \text{ times}}$.

In the field of rational numbers $\mathbb{Q}$ we have $n \cdot a \neq 0$ for all $a \in \mathbb{Q}$, $n \in \mathbb{Z}_{>0}$. On the other hand, if p is a prime number, then in the finite field F_p we have $p \cdot 1 = \underline{p} = 0$.

DEFINITION 1.1. Let k be a field. The *characteristic* of k, denoted $char(k)$, is the smallest $n \in \mathbb{Z}_{>0}$ with $n \cdot 1 = 0$ in k, if such an n exists, and otherwise $char(k) = 0$.

EXAMPLE 1.2. (1) $char(\mathbb{Q}) = char(\mathbb{R}) = char(\mathbb{C}) = 0$.
(2) $char(F_p) = p$.

DEFINITION 1.3. A subring $k \subset K$ of a field K is called a *subfield* if is is a field. In this case we call K a *field extension* of k. We write K/k is a field extension.

Let K/k and L/k be field extensions. A homomorphism $\varphi : K \to L$, with $\varphi(a) = a$ for all $a \in k$ is called a k-homomorphism. If φ is in addition an isomorphism, it is called a k-isomorphism. In this case we say that K and L are k-isomorphic.

REMARK 1.4. Let k be a subfield of a field K. Then $char(k) = char(K)$ (because the 1 of K is also the 1 of k).

If K/k is a field extension, then K is in particular a k-vector space. Its dimension will be called the degree of the field extension.

REMARK 1.5. Let K/k be a field extension. Then $(K, +)$ is an abelian group and the restriction of the multiplication in K defines a scalar multiplication

$$k \times K \to K;\ (a, x) \mapsto ax$$

so that the distributive laws $(a+b)x = ax+bx$, $a(x+y) = ax+ay$ and the associative law $(ab)x = a(bx)$ hold and furthermore $1x = x$. Thus K is a k-vector space.

DEFINITION 1.6. The *degree* $[K : k]$ of the field extension K/k is the dimension of K as a k-vector space (with $[K : k] = \infty$ if this is not a finite dimensional vector space). The field extension K/k is called *finite* if $[K : k] < \infty$.

REMARK 1.7. $[K : k] = 1$ if and only if $K = k$.

PROOF. $[K : k] = 1$ if and only if 1 is a k-basis of K if and only if $K = k \cdot 1 = k$. □

The degree of field extensions behaves multiplicatively.

DEFINITION 1.8. Let K/k be a field extension. Then a subfield $L \subset K$ with $k \subset L$ is called an intermediate field of K/k.

THEOREM 1.9. *(Degree theorem) If L is an intermediate field of a field extension K/k, then*

$$[K : k] = [K : L][L : k].$$

(This is with the convention $n\infty = \infty n = \infty\infty$ for all $n \in \mathbb{Z}_{>0}$.) In particular K/k is a finite extension if and only if K/L and L/k are finite extensions.

PROOF. If K/L or L/k is not finite, then K/k is trivially not finite.

So assume L/k and K/L are finite field extensions. Let $(x_1, \dots, x_n)$ be a basis of L over k and $(y_1, \dots, y_m)$ be a basis of K over L. We will show that $\{x_i y_j \mid i = 1, \dots, n,\ j = 1, \dots, m\}$ is a basis of K over k. They generate: If $y \in K$, then we can write

$$y = \sum_{j=1}^{m} b_j y_j, \quad b_j \in L,$$

and for all j we can write

$$b_j = \sum_{i=1}^{n} a_{ij} x_i, \quad a_{ij} \in k.$$

Thus we get $y = \sum_{i,j} a_{ij} x_i y_j$.

They are linearly independent: If

$$\sum_{i,j} a_{ij} x_i y_j = 0, \quad a_{ij} \in k,$$

then for all j we have $\sum_i a_{ij} x_i = 0$, because the y_j are linearly independent over L, and thus for all i, j we get $a_{ij} = 0$, because the x_i are linearly independent. □

COROLLARY 1.10. *If L is an intermediate field of a finite field extension K/k, then $[L : k]$ divides $[K : k]$. In particular if $[K : k]$ is a prime number, the only intermediate fields are k and K.*

Exercises.

(1) Let k be a field. Find all elements $a \in k$ with $a = a^{-1}$.
(2) Let F be a field with precisely 8 elements. Prove or disprove: The characteristic of F is 2.
(3) Let $a = 2^{\frac{1}{p}} \in \mathbb{C}$ for a prime number p. Show that $[\mathbb{Q}(a) : \mathbb{Q}] = p$ and the only intermediate fields of $\mathbb{Q}(a)/\mathbb{Q}$ are $\mathbb{Q}$ and $\mathbb{Q}(a)$.
(4) Let a be a positive rational number, which is not a square in $\mathbb{Q}$. Show that $[\mathbb{Q}(\sqrt[4]{a}) : \mathbb{Q}] = 4$.

2. Algebraic extensions and simple algebraic extensions

In this section we fix a field extension K/k. We are interested in elements of K which are zeros of polynomials in $k[x]$.

DEFINITION 2.1. An element $a \in K$ is called *algebraic* over k if there is a nonzero polynomial $f = \sum_{i=1}^{n} b_i x^i \in k[x]$ with $f(a) = \sum_i b_i a^i = 0$. $a \in K$ is called *transzendental* over k if it is not algebraic over k. K is called an *algebraic extension* of k if all elements of K are algebraic over k.

Thus an element $a \in K$ is transzendental if it is not the root of a nonzero polynomial in $k[x]$. In this course we will not deal with transzendental elements.

EXAMPLE 2.2. (1) $\sqrt{2} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$ because it is a zero of $x^2 - 2$.
(2) $i \in \mathbb{C}$ is algebraic over $\mathbb{Q}$, because it is a zero of $x^2 + 1$.
(3) It is known that e and π are transzendent over $\mathbb{Q}$.

DEFINITION 2.3. Let $a_1, \ldots, a_n \in K$. The *extension of k generated by $a_1, \ldots, a_n$* is the intersection of all subfields L of K which contain k and $a_1, \ldots, a_n$. It is denoted $k(a_1, \ldots, a_n)$. By definition $k(a_1, \ldots, a_n)$ is obviously a subfield of K containing k and $a_1, \ldots, a_n$.

Of special importance are the extensions generated by one algebraic element.

DEFINITION 2.4. If $a \in K$ is algebraic over k, then $k(a)$ is called a *simple algebraic extension* of k.

There are two reasons for the importance of simple algebraic extensions. First they can be understood very well. Secondly essentially all (in finite characteristic all) finite algebraic extensions are simple algebraic extensions (see the Theorem of the Primitive Element below). We start by giving an explicit description of a simple algebraic extension $k(a)$ in terms of the minimal polynomial of a, which is the unique irreducible monic polynomial $f \in k[x]$ with $f(a) = 0$.

DEFINITION 2.5. A polynomial $f \in k[x]$ is called monic, if its leading coefficient is 1. Note that for $f \in k[x] \setminus \{0\}$ there is a unique monic polynomial g with $\langle f \rangle = \langle g \rangle$. (If $b \in k^*$ is the leading coefficient of f, just put $g = f/b$).

Let $a \in K$ be algebraic over k. Let $ev_a : k[x] \to K, g \mapsto g(a)$ be the evaluation homomorphism. Then, by definition, a is algebraic if and only if $ker(ev_a) \neq 0$. As $k[x]$ is a PID, there is a unique monic polynomial $f_a \in k[x]$ with $ker(ev_a) = \langle f_a \rangle$. f_a is called the *minimal polynomial* of a over k.

PROPOSITION 2.6. *Let $a \in K$ be algebraic over k. The minimal polynomial f_a of a over k is the unique irreducible monic polynomial $f \in k[x]$ with $f(a) = 0$.*

PROOF. We want to show f_a is irreducible. Assume $f_a = gh$ with $g, h \in k[x]$. We have to show $g \in k^*$ or $h \in k^*$. We have $0 = f_a(a) = g(a)h(a)$, thus $g(a) = 0$ or $h(a) = 0$. If, say, $g(a) = 0$, then $g \in \langle f_a \rangle$, thus $g = lf_a$ for some $l \in k[x]$. Thus $f_a = gh = f_a lh$. Therefore $lh = 1$ and $l \in k^*$.

On the other hand let g be a monic irreducible polynomial with $g(a) = 0$. Then $g \in \langle f_a \rangle$, thus $g = lf_a$ for some $l \in k[x]$. As g is irreducible, we get $l \in k^*$. As both g and f_a are monic, this implies $l = 1$. □

EXAMPLE 2.7. For each prime number p and all $n \in \mathbb{Z}_{>1}$, we know that $x^n - p$ is irreducible in $\mathbb{Q}[x]$ by the criterion of Eisenstein. Thus $x^n - p$ is the minimal polynomial for $\sqrt[n]{q}$ over $\mathbb{Q}$.

Now let $k(a)/k$ be a simple algebraic extension. We can give a very explicit description of $k(a)$ in terms of the minimal polynomial of a over k.

THEOREM 2.8. *Let $a \in K$ be algebraic over k with minimal polynomial f_a, and $m := deg(f_a)$. Then*

(1) $k(a) \simeq k[x]/\langle f_a \rangle$,
(2) $[k(a) : k] = m$ *and* $(1, a, \dots a^{m-1})$ *is a basis of* $k(a)$ *over* k.

PROOF. (1) $ev_a : k[x] \to k(a)$; $g \mapsto g(a)$ is a ring homomorphism with kernel $\langle f_a \rangle$. Let L be the image of ev_a. Then by the homomorphism theorem $L \simeq k[x]/\langle f_a \rangle$. As f_a is irreducible and $k[x]$ is a PID, we know that $\langle f_a \rangle$ is a maximal ideal in $k[x]$ and $k[x]/\langle f_a \rangle$ is a field. Thus L is a subfield of $k(a)$. Furthermore $k \subset L$ as the image of the constant polynomials and $a = ev_a(x) \in L$. Thus $L = k(a)$.

(2) We have shown in (1) that $ev_a : k[x] \to k(a)$ is surjective, thus

$$k(a) = \big\{ g(a) \;\big|\; g \in k[x] \big\}.$$

We first show that $1, a, \dots, a^{m-1}$ generate $k(a)$ over k. Let $b \in k(a)$, then $b = g(a)$ for some $g \in k[x]$. If $deg(g) > deg(f_a)$, then division with rest gives

$$g = qf + r, \quad q, r \in k[x], \quad deg(r) < deg(f_a).$$

Then by $f_a(a) = 0$ we get $g(a) = r(a)$. Thus $b = r(a)$ for a polynomial $r = \sum_{i=0}^{m-1} b_i x^i$, i.e. $1, a, \dots, a^{m-1}$ generate $k(a)$. Assume $1, a, \dots, a^{m-1}$ are not linearly independent, then we can write $0 = \sum_{i=0}^{m-1} b_i a^i$, with $b_i \in k$ not all 0. Thus $h(a) = 0$ for $h = \sum_{i=0}^{m-1} b_i x^i$. On the other hand $h \in ker(ev_a) = \langle f_a \rangle$ implies $deg(h) \geq deg(f_a) = m$, a contradiction. □

REMARK 2.9. Let $k(b)/k$ be a simple algebraic extension of degree n and let $f = \sum_{i=0}^{n} a_i x^i$ be the minimal polynomial of b over k. Then by the Theorem $k(b)$ can be explicitly described as follows:

$$k(b) = \Big\{ \sum_{i=0}^{n-1} c_i b^i \;\Big|\; c_i \in k \Big\}.$$

Addition and multiplication are the usual ones as if these were polynomials in the indeterminate b together with the rule that we eliminate any power of b bigger than $n-1$ by $b^n = -\sum_{i=0}^{n-1} a_i b^i$.

EXAMPLE 2.10. The minimal polynomial of i over $\mathbb{R}$ is $x^2 + 1$. Thus $\mathbb{C} = \mathbb{R}(i) = \big\{ a + bi \;\big|\; a, b \in \mathbb{R} \big\}$, with addition and multiplication

$$(a + bi) + (c + di) = a + c + (b + d)i,$$
$$(a + bi)(c + di) = ac + (ad + bc)i + bdi^2 = ac - bd + (ad + bc)i.$$

We have seen that a simple algebraic extension is a finite extension. Now we want to see that conversely finite field extensions are algebraic extensions.

THEOREM 2.11. *Let K/k be a field extension. K/k is a finite extension if and only if K/k is algebraic and there are finitely many elements $a_1, \dots, a_n \in K$ such that $K = k(a_1, \dots, a_n)$.*

PROOF. "$\Longrightarrow$" Let $m := [K : k]$. Then for any $a \in K$ the elements $1, a, \dots, a^m$ are linearly dependent. Thus there exists a nonzero polynomial $f \in k[x]$ with $f(a) = 0$. If $a_1, \dots, a_n$ is a basis of K over k, then $K = k(a_1, \dots, a_n)$.

"$\Longleftarrow$" We prove this by induction on n, the case $n = 0$ being trivial. Assume that $K = k(a_1, \dots, a_{n+1})$ and that for $L := k(a_1, \dots, a_n)$ the degree $m := [L : k]$ is finite. Let g be the minimal polynomial of a_{n+1} over L. Then $[K : L] = deg(f)$ and $[K : k] = [K : L][L : k] = deg(f)m$ is finite. □

This result implies that the elements of K which are algebraic over k form an intermediate field of K/k.

COROLLARY 2.12. *Let* $L := \{a \in K \mid a \text{ is algebraic over } k\}$. *Then* L *is a subfield of* K/k *and* L/k *is an algebraic extension.*

PROOF. We know $k \subset L$. If $a, b \in L$, then $k(a, b)/k$ is a finite algebraic extension, and by the previous theorem also $a - b$, a/b lie in L. Thus L is a field, and by definition L/k is an algebraic extension. □

EXAMPLE 2.13. The set $\overline{\mathbb{Q}} := \{a \in \mathbb{C} \mid a \text{ is algebraic over } \mathbb{Q}\}$ is the field of algebraic numbers. It is an (infinite) algebraic extension of $\mathbb{Q}$.

We also get that the composition of algebraic extensions is an algebraic extension.

COROLLARY 2.14. *Let* K/L *and* L/k *be algebraic field extensions. Then* K/k *is an algebraic field extension.*

PROOF. Let $u \in K$. We need to show that u is algebraic over k. As u is algebraic over L, $g(u) = 0$ for a nonzero $g = \sum_{i=0}^{m} a_i x^i \in L[x]$. We put $M := k(a_0, \dots, a_m)$. Then M/k is a finite extension, and $M(u)/M$ is a finite extension. Thus $[M(u) : k] = [M(u) : M][M : k]$ is finite. Therefore u is algebraic over k. □

Extension of field homomorphims. Now we want to show the following: if $a, b \in K$ have the same minimal polynomial over k, then $k(a)$ is k-isomorphic to $k(b)$. We will prove something more general which will be a key ingredient in our development of Galois theory: If $\varphi : k \to k'$ is a field isomorphism which sends the minimal polynomial of $a \in K$ to that of $a' \in K'$, then there is a unique way to extend φ to an isomorphism $\Phi : k(a) \to k'(a')$ with $\Phi(a) = a'$.

DEFINITION 2.15. Let $\varphi : k \to k'$ a field isomorphism and let L/k, L'/k' be field extensions. An isomorphism $\Phi : L \to L'$ is called an *extension* of φ, if $\Phi|_k = \varphi$.

REMARK 2.16. Let $\varphi : k \to k'$ be an isomorphism of fields. Then φ defines an isomorphism

$$\varphi_* : k[x] \to k'[x]; f = \sum_i a_i x^i \mapsto \sum_i \varphi(a_i) x^i.$$

THEOREM 2.17. *Let $\varphi : k \to k'$ be a field isomorphism. Let L/k and L'/k' be field extensions. Let $a \in K$ be algebraic over k with minimal polynomial f_a. Let $a' \in K'$ be a zero of $\varphi_*(f_a)$.*

Then there is a unique extension $\Phi : k(a) \to k'(a')$ of φ with $\Phi(a) = a'$.

PROOF. (Uniqueness). Let $\Phi : k(a) \to k'(a')$ be an extension of φ with $\Phi(a) = a'$. Then $k(a) = \{g(a) | g \in k[x]\}$, and $k'(a') = \{g'(a') | g' \in k'[x]\}$ and for $g = \sum_i b_i x^i \in k[x]$, Φ is uniquely determined by:

$$\Phi(g(a)) = \sum_i \varphi(b_i)\Phi(a)^i = \sum_i \varphi(b_i)(a')^i = \varphi_*(g)(a').$$

(Existence) We define $\Phi : k(a) \to k'(a')$ by $\Phi(g(a)) := \varphi_*(g)(a')$. Then obviously Φ is a homomorphism. As φ_* is an isomorphism, we see that Φ is surjective. Finally assume $h(a) \in \ker(\Phi)$, for $h \in k[x]$. Then $\varphi_*(h)(a') = 0$. By definition $\varphi_*(f_a)$ is the minimal polynomial of a' over k'. Thus $\varphi_*(f_a)$ divides $\varphi_*(h)$ and thus f divides h and therefore $h(a) = 0$. This shows that Φ is injective, and thus an isomorphism. For $u \in k$ a we get $\Phi(u) = \varphi_*(u) = \varphi(u)$. And we get $\Phi(a) = \varphi_*(x)(a') = a'$. □

For the future the following corollary will be important:

COROLLARY 2.18. *Let K/k be a field extension and let $a, a' \in K$ be algebraic with the same minimal polynomial.*

Then there exists a unique k-isomorphism $\varphi : k(a) \to k(a')$ with $\varphi(a) = a'$.

Exercises.

(1) Let K/k be a field extension. Let $\alpha \in K$ be a zero of an irreducible polynomial $x^3 - ax + b \in k[x]$. Find explicitely the inverse of $1 + \alpha$ in $k(\alpha)$ in the form $u + v\alpha + w\alpha^2$ with $u, v, w \in k$.
(2) Let $n \in \mathbb{Z}_{\geq 1}$. Construct a field extension $K \subset \mathbb{C}$ of $\mathbb{Q}$ with $[K : \mathbb{Q}] = n$.
(3) Let p, q be distinct prime numbers and let $K := \mathbb{Q}(\sqrt{p}, \sqrt{q})$. Prove the following:
   (a) $[K : \mathbb{Q}] = 4$,
   (b) $K = \mathbb{Q}(\sqrt{p} + \sqrt{q})$,
   (c) The minimal polynomial of $\sqrt{p} + \sqrt{q}$ over $\mathbb{Q}$ is
   $$x^4 - 2(p + q)x^2 + (p - q)^2.$$
(4) Determine the minimal polynomial of $\sqrt{3} + \sqrt{5}$ over each of the following fields.
   (a) $\mathbb{Q}$.
   (b) $\mathbb{Q}(\sqrt{5})$.
   (c) $\mathbb{Q}(\sqrt{15})$.
(5) Let $a := \sqrt{2}, b := 3^{\frac{1}{3}} \in \mathbb{C}$. Show $[\mathbb{Q}(a, b) : \mathbb{Q}] = 6$.
(6) Determine $[\mathbb{Q}(\sqrt{2} + \sqrt{3}) : \mathbb{Q}]$.

(7) Assume $[k(a) : k] = n$, $[k(b) : k] = m$ and n, m are relatively prime. Show that $[k(a, b) : k] = nm$.
(8) Decide whether or not i is in the field
   (a) $\mathbb{Q}(\sqrt{-2})$.
   (b) $\mathbb{Q}(\sqrt[4]{-2})$.
   (c) $\mathbb{Q}(\alpha)$, where $\alpha^3 + \alpha + 1 = 0$.
(9) Let $\alpha := e^{2\pi i/7}$, $\beta := e^{2\pi i/5}$. Prove that $\beta \notin \mathbb{Q}(\alpha)$.
(10) Let α,β be complex numbers with $[\mathbb{Q}(\alpha) : \mathbb{Q}] = [\mathbb{Q}(\beta) : \mathbb{Q}] = 3$. Determine the possiblities for $[\mathbb{Q}(\alpha, \beta) : \mathbb{Q}]$.
(11) Let α, β be complex roots of irreducible polynomials $f, g \in \mathbb{Q}[x]$ respectively. Let $F := \mathbb{Q}(\alpha)$, $K := \mathbb{Q}(\beta)$. Show that f is irreducible in $K[x]$ if and only if g is irreducible in $F[x]$.
(12) Let $a \in \mathbb{C}$ be algebraic over $\mathbb{Q}$. We say that a is an algebraic integer if it satisfies a monic equation $a^n + b_{n-1}a^{n-1} + \ldots b_0 = 0$, with $b_0, \ldots, b_{n-1} \in \mathbb{Z}$.
   (a) Let $a \in \mathbb{C}$ be algebraic over $\mathbb{Q}$. Show there exists a positive integer m such that ma is an algebraic integer.
   (b) If $r \in \mathbb{Q}$ is an algebraic integer, then $r \in \mathbb{Z}$.
   (c) Let α be an algebraic integer satisfying $\alpha^3+\alpha+1 = 0$ and β an algebraic integer satisfying $\beta^2 + \beta - 3 = 0$. Show that $\alpha + \beta$ and $\alpha\beta$ are algebraic integers.
(13) Let k be a field and assume $x^n - a$ is irreducible over K and $m|n$. Let u be a zero of $x^n - a$ over an extension K/k. Show $[k(u^m) : k] = n/m$. Determine the minimal polynomial of u^m over k.
(14) Let k be a field, let a, b be algebraic over k with $[k(a) : k] = n$, $[k(b) : k] = m$. Show that $[k(a, b) : k] \leq nm$ with equality if n and m are relatively prime.

3. Algebraic closure

We will briefly without proofs introduce the algebraic closure of a field. The results will not be used in the rest of the course. A field K is called algebraically closed, if every nonconstant $f \in K[x]$ has a zero in K. An algebraic closure of a field k is an algebraically closed field which is an algebraic extension of k.

Definition 3.1. Let K be a field. K is called *algebraically closed* if the following equivalent statements hold.

(1) Every nonconstant polynomial $f \in K[x]$ has a zero in K.
(2) Every nonconstant polynomial $f \in K[x]$ splits into linear factors i.e. there exist $a_1, \ldots, a_n, b \in K$ such that

$$f = b(x - a_1) \cdot \ldots \cdot (x - a_n).$$

It is easy to see that the two statements are equivalent: (2)$\Longrightarrow$(1) is obvious. Now assume (1). Let $a \in K$ be a zero of a nonconstant polynomial $f \in K[x]$. Then we

can write $f = (x-a)g$ with $g \in K[x]$ and $deg(g) = deg(f) - 1$. The result follows by induction on $deg(f)$.

EXAMPLE 3.2. (1) $\mathbb{Q}$ is not algebraically closed, e.g. $x^2 - 2$ has no zero in $\mathbb{Q}$.
(2) $\mathbb{R}$ is not algebraically closed, e.g. $x^2 + 1$ has no zero in $\mathbb{R}$.
(3) $\mathbb{C}$ is algebraically closed. This is the *Fundamental Theorem of Algebra*. It is usually proved in Complex Analysis.
(4) $\overline{\mathbb{Q}}$ is algebraically closed.

DEFINITION 3.3. A field extension K/k is called an *algebraic closure* of k if K/k is an algebraic extension and K is algebraically closed.

With the Lemma of Zorn one shows that every field has up to isomorphism a unique algebraic closure. We will not go into the proof.

THEOREM 3.4. (1) *Every field k has an algebraic closure.*
(2) *If K, L are algebraic closures of k, then there exists a k-isomorphism $\Phi : K \to L$.*

EXAMPLE 3.5. (1) The algebraic closure of $\mathbb{R}$ is $\mathbb{C}$.
(2) The algebraic closure of $\mathbb{Q}$ is $\overline{\mathbb{Q}}$. Note that $\mathbb{C}$ is not an algebraic closure of $\mathbb{Q}$, because it contains transzendental elements, e.g. e, π.

Exercises.
(1) Show that $\overline{\mathbb{Q}}$ is algebraically closed.

4. Splitting fields

In this section we fix a field k. Given a nonconstant polynomial $f \in k[x]$ we want to find a finite field extension K/k such that k has a zero in K or even such that f splits into linear factors over K. First we show that for a nonconstant polynomial $f \in k[x]$ we can always find a root of f in a finite extension K of k.

THEOREM 4.1. *Let $f \in k[x]$ be irreducible. There exists a simple algebraic extension K/k with $[K : k] = deg(f)$, such that f has a zero in K.*

PROOF. This is a formal, almost tautological construction. We will use f to construct the field. As f is irreducible and $k[x]$ is a PID, we know that $K := k[x]/\langle f \rangle$ is field. Let $\pi : k[x] \to K$ be the canonical projection. $\pi|_k$ is injective and the image is a subfield of K (which we identify with k). Thus K/k is a field extension. We claim that f has a zero in K, namely the class $[x]$: Write $f := \sum_{i=0}^{n} a_i x^i$ with $a_i \in k$. Then

$$0 = [f] = \sum_{i=0}^{n} [a_i][x^i] = \sum_{i=0}^{n} a_i [x]^i = f([x]).$$

Thus $[x]$ is a zero of f in K, and $K = k([x])$.

Finally we determine $[K : k]$. By dividing f by its leading coefficient we can assume that f is monic. So f is an irreducible monic polynomial with $f([x]) = 0$, that is f is the minimal polynomial of $[x]$ over k. Thus $[K : k] = deg(f)$. □

DEFINITION 4.2. In the situation of the Theorem we say that K is obtained from k *by formally adjoining a root of* f.

COROLLARY 4.3. *Let $f \in k[x]$ be a polynomial of degree $n > 0$. Then there exists a field extension K/k with $[K : k] \leq n$ such that f has a zero in K.*

PROOF. Just apply the previous theorem to any irreducible factor of f. □

Given a polynomial $f \in k[x]$ we can adjoin successively more roots until f splits into linear factors. The smallest such extension of k will be called a splitting field of f over k. It is uniquely determined up to isomorphism.

DEFINITION 4.4. Let $f \in k[x]$ by a polynomial of degree $n > 0$. A finite extension K/k is called a *splitting field* of f over k if

(1) f splits over K into linear factors, i.e. there exist $a_1, \ldots, a_n, b \in K$ such that $f = b(x - a_1) \cdot \ldots \cdot (x - a_n)$.
(2) f does not split over any intermediate field $K \supsetneq L \supset k$.

It is easy to see that splitting fields of polynomials always exist.

COROLLARY 4.5. *Let $f \in k[x]$ be monic.*

(1) *If L/k is an extension of k over which f splits into linear factors $x - a_1, \ldots, x - a_n$ then $k(a_1, \ldots, a_n)$ is a splitting field of f over k.*
(2) *There exists a splitting field K of f over k with $[K : k] \leq n!$.*
(3) *Let K be a splitting field of f over k, and let L be an intermediate field. Then K is also a splitting field of f over L.*

PROOF. (1) Let L/k be a field extension, such that over L we have $f = (x - a_1) \ldots (x - a_n)$ with $a_i \in L$. We claim that $F := k(a_1, \ldots, a_n)$ is a splitting field of f. Obviously f splits over F into linear factors. Assume $L_1 \subset F$ is a subfield over which f factors into linear factors $f = (x - c_1) \ldots (x - c_n)$. Then for all i we have $0 = f(a_i) = (a_i - c_1) \ldots (a_i - c_n)$. Therefore $a_i = c_j$ for some j. Thus we see that $a_i \in L_1$ for all i, i.e. $F := k(a_1, \ldots, a_n) \subset L_1$. Therefore $F = L_1$.

(2) By the previous corollary we find an extension K_1/k with $[K_1 : k] \leq n$ such that f has a root $a_1 \in K_1$. Then $f = (x - a_1)g$ with $deg(g) \leq n - 1$. By induction there is a finite field extension K/k of degree $\leq n!$ such that f splits over K into linear factors $x - a_1, \ldots, x - a_n$. By (1) $k(a_1, \ldots, a_n)$ is a splitting field of f over k.

(3) f splits over K into linear factors, and there is no intermediate field between L and K where it splits. □

EXAMPLE 4.6. (1) $\mathbb{C}$ is a splitting field of $x^2 + 1$ over $\mathbb{R}[x]$ and $\mathbb{Q}[\sqrt{2}]$ is a splitting field of $x^2 - 2$ over $\mathbb{Q}$.

(2) More generally let $f \in k[x]$ be an irreducible polynomial of degree 2 and K/k is an extension of degree 2 such that f has a zero in K. Then K is a splitting field of f over k.
(3) (Splitting field of x^4+1 over $\mathbb{Q}$). Let α be a root of x^4+1 in an extension of $\mathbb{Q}$. Then also $-\alpha$, $\frac{1}{\alpha}$ and $-\frac{1}{\alpha}$ are roots of x^4+1. These roots are distinct: $\alpha \neq -\alpha$ because $\alpha \neq 0$ and if $\alpha = \pm\frac{1}{\alpha}$, then $\alpha^2 = \pm 1$; thus $\alpha^4 + 1 = 2 \neq 0$. Thus over $\mathbb{Q}(\alpha)$ we get a splitting
$$x^4 - 1 = (x-\alpha)(x+\alpha)(x-\frac{1}{\alpha})(x+\frac{1}{\alpha}).$$
Therefore $\mathbb{Q}(\alpha)$ is the splitting field of x^4+1.
(4) (Splitting field of x^3-2). Over the complex numbers we have
$$x^3 - 2 = (x - \sqrt[3]{2})(x - e^{2\pi i/3}\sqrt[3]{2})(x - e^{4\pi i/3}\sqrt[3]{2}).$$
Clearly x^3-2 splits over $\mathbb{Q}(\sqrt[3]{2}, e^{2\pi i/3})$, but it does not split over $\mathbb{Q}(\sqrt[3]{2})$ or $\mathbb{Q}(e^{2\pi i/3})$, because $\sqrt[3]{2}$ is real and $\mathbb{Q}(e^{2\pi i/3})$ does not contain $\sqrt[3]{2}$. $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$ and $[\mathbb{Q}(e^{2\pi i/3}) : \mathbb{Q}] = 2$. Thus $[\mathbb{Q}(\sqrt[3]{2}, e^{2\pi i/3}) : \mathbb{Q}] = 6$.

Extension of field isomorphisms. The aim of Galois theory will be to understand field extensions K/k in terms of the k-automorphisms of K. Thus it is important to study how field isomorphisms extend.

Similarly to the case of simple algebraic extensions, we want to show that a field isomorphism $\varphi : k \to k'$ can be extended to an isomorphism of splitting fields $\Phi : K \to K'$ were K is a splitting field of a polynomial $f \in k[x]$ and K' is a splitting field of $\varphi_*(f)$. This in particular shows that the splitting field of a polynomial $f \in k[x]$ is uniquely determined up to isomorphism.

THEOREM 4.7. *Let $\varphi : k \to k'$ be an isomorphism of fields. Let $f \in k[x]$ be a nonzero polynomial and $\widehat{f} := \varphi_*(f)$. Let K be the splitting field of f over k and K' the splitting field of $\widehat{f}$ over k'. Then there is an isomorphism $\Phi : K \to K'$ with $\Phi|_k = \varphi$.*

In particular if K, K' are splitting fields of f over k then there is a k-isomorphism $\Phi : K \to K'$.

PROOF. We use induction over the degree $[K : k]$. If $[K : k] = 1$, then $K = k$ and $K' = k'$, so there is nothing to show.

If $[K : k] > 1$, then f contains an irreducible factor g of degree $deg(g) > 1$ and similarly $\widehat{g} = \varphi_*(g)$ is an irreducible factor of $\widehat{f}$ of degree > 1. Let a be a zero of g in K and a' a zero of $\widehat{g}$ in K'. Then by the previous theorem there is an isomorphism $\varphi' : k(a) \to k'(a')$ with $\varphi'|_k = \varphi$ and $[K : k(a)] < [K : k]$.

On the other hand K is the splitting field of f over $k(a)$ and K' is the splitting field of $\widehat{f}$ over $k'(a')$. Thus by induction there is an isomorphism $\Phi : K \to K'$ with $\Phi|_{k(a)} = \varphi'$ and thus $\Phi|_k = \varphi$. □

REMARK 4.8. Note that the result that we have proven here is much weaker than the result that we had earlier for simple algebraic extensions. While in the case of simple algebraic extensions we had existence and *uniqueness* of the extension $k(a) \to k'(a')$ (for any zero a of the minimal polynomial f of a and a' of $\varphi_*(f)$), now we only have the existence. Thus we do not know how many extensions there are. This is the main difficulty when doing Galois theory.

Exercises.

(1) Let $f = x^4 + x^2 + 1 \in \mathbb{Q}[x]$. Let $\omega := \frac{-1+\sqrt{3}i}{2}$. Show that $\mathbb{Q}(\omega)$ is the splitting field of f over $\mathbb{Q}$ and that $[\mathbb{Q}(\omega) : \mathbb{Q}] = 2$.
(2) Determine the degrees of the splitting fields of the following polynomials over $\mathbb{Q}$.
   (a) $x^4 + 1$.
   (b) $x^6 + 1$.
   (c) $x^4 - 2$.
   (d) $x^6 + x^3 + 1$.
(3) Let p be a prime number. Show that the degree of the splitting field of $x^p - 1$ over $\mathbb{Q}$ is $p - 1$.
(4) Let $f := x^3 + ax + b \in \mathbb{Q}[x]$. Find necessary and sufficient conditions on a and b so that the degree of the splitting field of f over $\mathbb{Q}$ is 3.
(5) Let k be a field, and let K/k be a field extension. Let $f \in k[x]$ and let φ be an automorphism of K with $\varphi|_k = id_k$. Show that φ maps a zero of $f \in K$ to a zero of $f \in K$.
(6) Show that $\mathbb{Q}(\sqrt[3]{2})$ has no automorphism except for the identy.
(7) Let $f \in \mathbb{R}[x]$. If $\alpha \in \mathbb{C}$ is a zero if f, then also the complex conjugate $\overline{\alpha}$.
(8) Let k be a field and let $f \in k[x]$ be irreducible of degree 6. Let K/k be an extension of degree 2. Prove or disprove: Either f is irreducible over K or f is the product of two irreducible polynomials of degree 3 over K.

5. Normal extensions

Now we come to two very important properties of field extensions: normal extensions and separable extensions. If a field extension is both normal and separable, then it will be called later a Galois extension, and the Galois extensions are those that we will mostly want to study. Both the definition of normal extension and of separable extension are not very intuitive, however we will see that the normal extensions are precisely the splitting fields and in characteristic 0 all field extensions are separable.

DEFINITION 5.1. A field extension K/k is called *normal* if it is an algebraic extension and every irreducible polynomial $f \in k[x]$ that has a zero in K splits over K into linear factors.

EXAMPLE 5.2. In the above example $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is not a normal extension, because $x^3 - 2$ has a root in $\mathbb{Q}(\sqrt[3]{2})$, but does not split into linear factors.

PROPOSITION 5.3. *A finite extension K/k is normal if and only if K is the splitting field of a polynomial in $k[x]$.*

PROOF. "$\Longleftarrow$" Let K be a splitting field of a polynomial $f \in k[x]$. Let $g \in k[x]$ be an irreducible polynomial with a zero $\alpha \in K$. We have to show that g splits over K. Let β be another zero of g in an extension of K. We will show $\beta \in K$. Then the result follows by induction.

Since g is irreducible, we know that there is an k-isomorphism $\varphi : k(\alpha) \to k(\beta)$. K is also a splitting field of f over $k(\alpha)$, and $K(\beta)$ is a splitting field of f over $k(\beta)$. Thus we get an isomorphism $\Phi : K \to K(\beta)$ with $\Phi|_{k(\alpha)} = \varphi$ and thus $\Phi|_k = id$. Thus $[K : k] = [K(\beta) : k]$ and therefore $[K(\beta) : K] = 1$, which implies $K(\beta) = K$. Thus $\beta \in K$.

"$\Longrightarrow$" Let K/k be a finite normal extension. We can write $K = k(a_1, \dots, a_n)$ with e.g. the a_i a basis of K over k. Let f_i be the minimal polynomial of a_i over k. Then because K/k is normal, each f_i splits over $k(a_1, \dots, a_n)$ into linear factors, and thus $f = f_1 \dots f_n$ does. As $K = k(a_1, \dots, a_n)$ and $a_1, \dots, a_n$ are the zeros of f, it is the splitting field of f. □

EXAMPLE 5.4. (1) $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ is a normal extension.
(2) $\mathbb{Q}(\sqrt[3]{2}, e^{2\pi i/3})$ is a normal extension.

6. Separable extensions

Let k be a field. Let $f \in k[x]$ be a polynomial and let $f = b(x - a_1) \dots (x - a_n)$ be a splitting of f into linear factors over some extension of L/k. It can happen that some of the a_i coincide (e.g. $a_1 = a_2 = a$). We say in this case that a is a multiple root of f. To avoid problems coming from multiple roots, the concept of separable extension is introduced. In characteristic zero all algebraic extensions are separable.

DEFINITION 6.1. Let $f \in k[x]$ and let

$$f = b(x - a_1)^{m_1} \dots (x - a_l)^{m_l}, \quad m_i > 0, \quad a_1, \dots, a_l \in K \text{ distinct}$$

be the splitting of f into linear factors over an extension K/k. If $m_i = 1$, then a_i is called a *simple root* in K. If $m_i > 1$, then a_i is called a *multiple root* of order m_i.

DEFINITION 6.2. An algebraic field extension K/k is called *separable* if all irreducible polynomials $f \in k[x]$ with a zero in K have only simple roots in their splitting fields. Equivalently the minimal polynomial of any $a \in K$ has only simple roots in its splitting field. If every algebraic extension K/k is separable over k, then k is called *perfect*.

We need a criterion for $f \in k[x]$ to have multiple roots in its splitting field. Like in calculus an element $a \in K$ is a multiple root of f if and only if it is both a root of f and of the derivative f'. Here the derivative is done by formally applying the rules that one knows from calculus for polynomials in $\mathbb{R}[x]$.

DEFINITION 6.3. Let $f = \sum_{i=0}^{n} a_i x^i \in k[x]$. The *derivative* of f is $f' := \sum_{i=1}^{n} i a_i x^{i-1}$.

REMARK 6.4. It is straightforward to check and left as an exercise that the usual rules for differentiation hold:

(1) $(af + bg)' = af' + bg'$ for $a, b \in k$
(2) $(fg)' = f'g + fg'$.

LEMMA 6.5. *Let $f \in k[x]$ be a nonconstant polynomial and let K/k be a splitting field of f, let $a \in K$ be a zero of f. Then a is a multiple root of f if and only if $f'(a) = 0$.*

PROOF. Let $r \geq 1$ be the multiplicity of the root a. Then we can write $f = (x-a)^r g$ with $g \in K[x]$ and $g(a) \neq 0$ and we get

$$f' = r(x-a)^{r-1} g + (x-a)^r g'$$

and the claim follows. □

THEOREM 6.6. *Let $f \in k[x]$ be an irreducible polynomial. f has no multiple roots in its splitting field over k if and only if $f' \neq 0$.*

PROOF. Let K/k be the splitting field of f. If $f' = 0$, then by the above every root of f is a multiple root. Let $a \in K$ be a multiple root of f. Then $f(a) = f'(a) = 0$. As f is irreducible, it is up to a constant multiple the minimal polynomial of a. Thus f' is divisible by f. If $f' \neq 0$, then $deg(f') < deg(f)$, so we get a contradiction. □

In calculus we know that a polynomial $f \in \mathbb{R}[x]$ can only have 0 derivative if it is constant, in particular irreducible polynomials $f \in \mathbb{R}[x]$ can never have derivative 0. It is easy to see that the same holds if the characteristic of k is 0. Thus we obtain:

COROLLARY 6.7. *Every field of characteristic 0 is perfect.*

PROOF. It follows from the definition that for a nonconstant polynomial $f \in k[x]$ over a field k of characteristic 0 the derivative f' is not 0. □

We finish this section with a proof of the important theorem of the primitive element. It says that every finite separable extension is a simple extension. In particular in characteristic 0 all finite extensions are simple extensions. This is very useful because we understand simple extensions much better.

THEOREM 6.8. *(Theorem of the primitive element). Let K be a finite separable field extension of k. Then there exists an element $a \in K$ with $K = k(a)$.*

PROOF. In the proof we assume for simplicity that k is infinite. (This is for instance true of k has characteristic 0. The theorem also holds for finite fields, but one needs some additional arguments.) The proof is slightly subtle.

Since $[K : k]$ is finite, we have $K = k(a_1, \ldots, a_n)$ for some $a_i \in K$. We make induction on n. If $n > 2$, then by induction we have $k(a_1, \ldots, a_{n-1}) = k(a)$ for some

$a \in K$ and $K = k(a, a_n)$. Thus we can assume $n = 2$ and write $K = k(a, b)$. We will show that $K = k(c)$ for a general linear combination $c = a + zb$ with $z \in k$.

Let f and g be the minimal polynomials of a and b over k respectively. Let L/K be a field where f and g split into linear factors. Let

$$a = x_1, x_2, \dots, x_n \text{ roots of } f;$$
$$b = y_1, y_2, \dots, y_m \text{ roots of } g.$$

Then, as K/k is separable, $b \neq y_j$ for all $j \neq 1$. It follows that for $i = 1, \dots, n$, $j = 2, \dots, m$, the element $z_{ij} := \frac{x_i - a}{b - y_j}$ is the only element of L with $a + z_{ij}b = x_i + z_{ij}y_j$. Since k is infinite, we can choose an element $z \in k$, different from all z_{ij}, thus $a + zb \neq x_i + zy_j$ unless $i = j = 1$.

Put $c = a + zb$. Then obviously $k(c) \subset k(a, b)$. We want to show $k(a, b) \subset k(c)$. We define $h \in k(c)[x]$ by $h(x) := f(c - zx)$. Then

$$h(b) = f(c - zb) = f(a) = 0.$$

As b is a zero of g in L we get that $(x - b)$ is a common factor of h and g in $L[x]$. We want to show that $x - b$ is a greatest common divisor of h and g in $L[x]$. As g splits over L into linear factors, the greatest common divisor must be a product of some linear factors of g. For $y_j \neq b$ another root of g, we get $h(y_j) = f(c - zy_j) \neq 0$ because by our choice of z, $c - zy_j \neq a_i$ for all roots a_i of f. Thus $(x - y_j)$ is not a factor of h. Thus $(x - b)$ is the greatest common divisor of h and g in $L[x]$.

However by the Euclidean algorithm a monic greatest common divisor of h, g lies in $k(c)[x]$, thus $x - b \in k(c)[x]$, i.e. $b \in k(c)$ and $a = c + zb \in k(c)$. Thus $k(a, b) \subset k(c)$. □

Exercises.

(1) Find a primitive element for the extension $\mathbb{Q}(\sqrt[3]{2}, e^{2\pi i/3})/\mathbb{Q}$.
(2) Let k be a field of characteristic $p \neq 0$ and let $f \in k[x]$ with $f' = 0$. Show that $f = g(x^p)$ for some polynomial $g \in k[x]$.
(3) Let $K/L/k$ be field extensions. Show that if K is normal over L and L is normal over k then K need not be normal over k. (Hint: consider $\mathbb{Q}(\sqrt[4]{2})/\mathbb{Q}(\sqrt{2})/\mathbb{Q}$).
(4) Let K/k be a field extension with $[K : k] = 2$. Show that K/k is normal.
(5) Let k be a field of characteristic 0. Let $f \in k[x]$. Let $g \in k[x]$ be an irreducible polynomial that divides f and f'. Show g^2 divides f.
(6) For which fields and which primes p does $x^p - x$ have a multiple root?

7. Finite fields

Before going on with the general theory we briefly want to study finite fields. For a prime number p we already know the finite field $F_p = \mathbb{Z}/p\mathbb{Z}$ with p elements. One might think that these are the only finite fields, but this is not true: For every prime power $q = p^n$ there is up to isomorphism a unique field with q elements, the splitting

field of $x^q - x$ over F_p. First we show that for any prime power $q = p^n$ there is a field with q elements. For the rest of this section we fix a prime number p, a positive integer n and put $q = p^n$.

LEMMA 7.1. *Let F be a field of characteristic p. Then $x^q - x \in F[x]$ has precisely q simple roots in its splitting field.*

PROOF. Let $a \in F$ be a multiple root. Then a is a zero of $x^q - x$ and of $(x^q - x)' = qx^{q-1} - 1 = -1$, which is impossible. □

PROPOSITION 7.2. *There exists a field with q elements.*

PROOF. Let K be the splitting field of $x^q - x$ over F_p. Let

$$F := \{a \in K \mid (a^q - a) = 0\},$$

be the set of roots of $(x^q - x)$. By the previous lemma F has q elements. We claim that F is a field. If $a, b \in F$, then $(ab)^q = a^q b^q = ab$, thus $ab \in F$, $(1/a)^q = 1/a^q = 1/a$, so $1/a \in F$ and $(a \pm b)^q = a^q \pm b^q = (a \pm b)$. Thus $a \pm b \in F$. Thus F is a field with q elements. □

The argument below will show that indeed $F = K$. Now we want to show that these are up to isomorphism all finite fields.

REMARK 7.3. Let F be a finite field of characteristic p. Then F contains $\{n \cdot 1 | n \in \mathbb{Z}\} \simeq F_p$. Thus F is a finite extension of F_p. Putting $n = [F : F_p]$, then F has p^n elements.

PROPOSITION 7.4. *If F is a field with $q := p^n$ elements, then F is a splitting field of the polynomial $x^q - x$ over F_p.*

PROOF. First we want to see that every element $a \in F$ satisfies $a^q - a = 0$. If $a = 0$, this is obvious. $(F \setminus \{0\}, \cdot)$ is a group of order $q - 1$. By a well known result in group theory (small theorem of Fermat) for every element g in a finite group G of order k we have $g^k = e$ where e is the neutral element of G. Thus we get $a^{q-1} = 1$ and thus $a^q - a = 0$.

As $x^q - x$ has degree q it can have at most q roots in F; on the other hand the elements of F are q distinct roots. Thus $x^q - x$ splits over F into linear factors. F is the splitting field of $x^q - x$, because it consists only of roots of $x^q - x$. □

Putting these two proposition together and recalling that the splitting field of a polynomial over a given field is unique up to isomorphism we get a complete classification of the finite fields.

THEOREM 7.5. *A finite field with n elements exists only if n is a prime power. For each prime power $q = p^m$ there is up to isomorphism a unique field with q elements: the splitting field of $x^q - x$ over F_p.*

Exercises.

(1) Identify the additive group of F_4.
(2) Determine the number of irreducible polynomials of degree 3 over F_2.
(3) Determine all polynomials $f \in F_q[x]$ such that $f(\alpha) = 0$ for all $\alpha \in F_q$.
(4) Prove that every element in F_p has exactly one p-th root.

8. Galois groups

The aim of Galois theory is to study fields via their automorphism groups. More precisely one studies finite field extensions K/k via the Galois group $Gal(K/k)$ of k-automorphisms of K. Very much information about the field extension is encoded in the Galois group. It has been said that modern number theory is nothing else then the study of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$.

DEFINITION 8.1. Let K/k be a field extension. The Galois group $Gal(K/k)$ of K over k is the set of k-automorphisms $\varphi : K \to K$. Obviously the composition of two k automorphisms is a k-automorphism, and we see that $Gal(K/k)$ is a group.

EXAMPLE 8.2. Complex conjugation $\iota : \mathbb{C} \to \mathbb{C}, a + bi \mapsto a - bi$, is an element of $Gal(\mathbb{C}/\mathbb{R})$. We will see that $Gal(\mathbb{C}/\mathbb{R}) = \{id, \iota\}$.

At first look the Galois group looks like something very abstract that is very difficult to understand. However for a simple algebraic extension $k(a)/k$, we can identify $Gal(k(a)/k)$ with a subgroup of the group of permutations of the roots of the minimal polynomial of a.

NOTATION 8.3. For a finite set M we denote by $S(M) := \{\sigma : M \to M \text{ bijection}\}$ the set of permutations of M. Then $S(M)$ is isomorphic to the symmetric group $\mathcal{S}_n$ where $n = |M|$ is the number of elements of M.

Recall that an action of a group G on a set M is *simply transitive* if for all $m_1, m_2 \in M$ there is a unique $g \in G$ with $g(m_1) = m_2$. Then obviously $|G| = |M|$.

THEOREM 8.4. *Let $k(a)/k$ be a simple algebraic extension of degree n. Let f be the minimal polynomial of a over k and let $R := \{b \in k(a) \mid f(b) = 0\}$ be the set of its roots.*

Then $Gal(k(a)/k)$ acts simply transitively on R. Thus $Gal(k(a)/k)$ is isomorphic to a subgroup of $S(R)$ of order $|R| \leq n$.

PROOF. Let $\varphi \in Gal(k(a)/k)$. If $g = \sum_i a_i x^i \in k[x]$ and $b \in k(a)$, then

$$\varphi(g(b)) = \sum_i a_i \varphi(b)^i = g(\varphi(b)).$$

Thus if b is a root of f, then $0 = f(b) = \varphi(f(b)) = f(\varphi(b))$, thus $\varphi(b)$ is also a root of f. (This elementary argument will be used very often in the future). So we see

that $\varphi|_R$ maps R into itself. As φ is injective, also $\varphi|_R$ is injective, and as R is finite it follows that $\varphi|_R$ is a bijection of R to itself. Thus we get a map

$$res_R : Gal(k(a)/k) \to S(R), \varphi \mapsto \varphi|_R.$$

Obviously $(\varphi \circ \psi)|_R = \varphi|_R \circ \psi|_R$, i.e. $res_R : Gal(k(a)/k) \to S(R)$ is a group homomorphism.

We have seen that given any two roots b_1, b_2 of f, there exists a unique element $\varphi \in Gal(k(a)/k)$ with $\varphi(b_1) = b_2$. Thus $Gal(k(a)/k)$ acts simply transitively on R, in particular res_R is injective. Therefore $Gal(k(a)/k)$ is isomorphic to a subgroup of $S(R)$. As the action is simply transitive we get $|Gal(k(a)/k)| = |R|$. Finally $|R| \leq deg(f) = n$. □

EXAMPLE 8.5. (1) $Gal(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$ consists of the two elements id and $a + b\sqrt{2} \mapsto a - \sqrt{2}$: The minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$ which has the two roots $\pm\sqrt{2}$.

(2) $Gal(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}) = id$: The minimal polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$, which has $\sqrt[3]{2}$ as only root in $\mathbb{Q}(\sqrt[3]{2})$. Thus $Gal(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q})$ contains only one element.

The subject of Galois theory are Galois extensions, which are finite separable normal extensions.

DEFINITION 8.6. A finite field extension K/k is called a *Galois extension* if it is separated and normal. (Note that in characteristic 0 separatedness is automatic).

COROLLARY 8.7. *Let K/k be a Galois extension of degree n. Then $Gal(K/k)$ is isomorphic to a subgroup of $\mathcal{S}_n$ which acts simply transitively and $|Gal(K/k)| = n$.*

PROOF. By the Theorem of the Primitive Element $K = k(a)$ for some $a \in K$, and as K/k is separated and normal the minimal polynomial of f has n distinct roots in K. □

PROPOSITION 8.8. *Let K/k be a Galois extension and $a, b \in K$. There exists an element $\varphi \in Gal(K/k)$ with $\varphi(a) = b$ if and only if a and b have the same minimal polynomial over k.*

PROOF. If a, b have the same minimal polynomial, then there is a unique k-isomorphism $\psi : k(a) \to k(b)$ with $\psi(a) = b$. K is the splitting field of a polynomial $f \in k[x]$ over k and thus also over $k(a)$. Thus by the extension of isomorphisms to splitting fields ψ can be extended to an element $\varphi \in Gal(K/k)$.

For the converse let $b = \varphi(a)$, with $a, b \in K$, $\varphi \in Gal(K/k)$. Let f and g be the minimal polynomials of a, b over k. Then $0 = \varphi(f(a)) = f(\varphi(a)) = f(b)$, thus g divides f, as g is the minimal polynomial of b. Similarly $f|g$. Therefore we get $f = g$, as f, g are monic. □

We will see the power of this result in the proof of the next theorem. A very important result for Galois theory is that k is precisely the set of elements of a Galois extension K/k fixed by all elements of $Gal(K/k)$. This is half of the proof of the Main Theorem of Galois theory, which relates subgroups of $Gal(K/k)$ to intermediate fields $K/L/k$

THEOREM 8.9. *Let K/k be a Galois extension. Then $k = \{a \in K \mid \varphi(a) = a \text{ for all } \varphi \in Gal(K/k)\}$.*

PROOF. "$\subset$" is trivial. "$\supset$" Assume $\varphi(a) = a$ for all $\varphi \in Gal(K/k)$. Let f be the minimal polynomial of a over k. Since K/k is normal, f splits over K. Let b be a root of f. Then there exists $\varphi \in Gal(K/k)$ with $\varphi(a) = b$. Thus $a = b$ by our assumption, i.e all the roots of f are equal. But f does not have multiple roots because K/k is separable. Thus $f = x - a \in k[x]$ and thus $a \in k$. □

We will usually study field extensions given as the splitting field of some polynomial. Thus we define the Galois group of a polynomial $f \in k[x]$ as the Galois group of its splitting field K over k. In this case $Gal(f)$ can be identified with a subgroup of the permutations of the roots of f in K, and this is usually also the best way to look at it.

DEFINITION 8.10. Let $f \in k[x]$ be a nonconstant polynomial. Let K be the splitting field of f over K (unique up to isomorphism which is the identity on k). The *Galois group* of f is $Gal(f) := Gal(K/k)$.

PROPOSITION 8.11. *Let $f \in k[x]$ be a nonconstant polynomial of degree n. Let R be the set of roots of f in the splitting field K of f.*

1. *Then $Gal(f)$ is isomorphic to a subgroup of $S(R)$ and $|Gal(f)|$ divides $n!$.*
2. *If all roots of f in K are simple, then f is irreducible if and only if $Gal(f)$ acts transitively on R.*

PROOF. (1) We have $Gal(f) = Gal(K/k)$. If a is a root of f in K, then $f(\varphi(a)) = \varphi(f(a)) = 0$ for all $\varphi \in Gal(K/k)$. Thus restriction to R defines a group homomorphism $Gal(K/k) \to S(R)$. If $\varphi|_R = id$, then $\varphi = id$, because $K = k(R)$. Hence $Gal(f)$ is isomorphic to a subgroup of $S(R)$. As $|R| \leq n$, we get that $|Gal(f)|$ divides $|R|!$ which divides $n!$.

(2) Now assume that f has only simple roots in K. Let f be irreducible, $a, b \in R$. Then f is the minimal polynomial of a and b over k. There exists an k-isomorphism $\psi : k(a) \to k(b)$ with $\psi(a) = b$. K is the splitting field of f both over $k(a)$ and $k(b)$. Thus by the extension of isomorphisms to splitting fields there exists an extension $\varphi \in Gal(K/k)$ of ψ. Thus $Gal(f)$ acts transitively on R.

Conversely assume $Gal(f)$ acts transitively on R, but f is reducible. Let $g_1, g_2 \in k[x]$ be two different irreducible factors of f. Let a_1, a_2 be roots of g_1, g_2 in K respectively. As $Gal(K/k)$ acts transitively, there exists $\varphi \in Gal(f)$ with $\varphi(a_1) = a_2$ But

we have

$$0 = \varphi(g_1(a_1)) = g_1(\varphi(a_1)) = g_1(a_2).$$

Thus g_2, which is the minimal polynomial of a_2 divides g_1. By irreducibility $g_1 = g_2$ (up to constant factors). Thus $g_1^2 | f$ and thus f cannot have only simple roots. □

REMARK 8.12. There is one point that you may find confusing: Let K/k be a Galois extension, which is the splitting field of a polynomial $f \in k[x]$ of degree n. Let $N := [K : k]$. Then we know that $Gal(f) = Gal(K/k)$ has N elements. $Gal(f)$ is in two different ways a subgroup of a symmetric group:

(1) As we just saw, if R is the set of roots of f in K, then $Gal(f)$ is via $\varphi \mapsto \varphi|_R$ a subgroup of $S(R)$, the group of permutations of R. This is how one should always view $Gal(f)$. Note that $|R| = n$.
(2) As K/k is a finite algebraic extension there is a primitive element $a \in K$, with $K = k(a)$. Let g be the minimal polynomial of a over k. Then g has degree N. Let Σ be the set of roots of g in K. Thus also $|\Sigma| = N$. Then $Gal(f) = Gal(K/k)$ is subgroup acting simply transitively of $S(\Sigma)$. This second viewpoint is useful for some proofs, but it is not the way to look at the Galois group $Gal(f)$ in examples, because usually we do not know how to find the primitive element a.

EXAMPLE 8.13.
(1) Let $f = x^3 - 2 \in \mathbb{Q}[x]$. We claim that $Gal(f) = \mathcal{S}_3$.

f is irreducible. Let $K/\mathbb{Q}$ be the splitting field, which is a Galois extension. Then $K = \mathbb{Q}[2^{\frac{1}{3}}, e^{2\pi i/3}]$, and $[K : \mathbb{Q}] = 6$. Thus $|Gal(f)| = 6$. On the other hand $Gal(f)$ is a subgroup of $\mathcal{S}_3$, thus $Gal(f) = \mathcal{S}_3$.

(2) Let $f = x^4 - 1 \in \mathbb{Q}[x]$. We claim that $Gal(f)$ is F_3^* which is also the cyclic group with 3 elements.

Let $a = e^{2\pi i/4}$. Then the roots of $f \in \mathbb{C}$ are $1, a, a^2, a^3$, thus the splitting field of f is $\mathbb{Q}(a)$. We have $f = (x-1)(1+x+x^2+x^3)$ and the second factor is irreducible. Thus the minimal polynomial of a is $1 + x + x^2 + x^3$. Therefore $|Gal(f)| = [\mathbb{Q}(a) : \mathbb{Q}] = 3$. We define a group homomorphism $\theta : F_3^* \to Gal(f), i \mapsto \theta(i)$, where $\theta(i)$ is the element of $Gal(f)$ defined by $\theta(i)(a) = a^i$. Obviously θ is injective. As both groups have the same number of elements, they are isomorphic.

(3) Let $f = x^4 + 1 \in \mathbb{Q}[x]$. We claim that $Gal(f) = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$.

If $\alpha = e^{2\pi i/8}$ is a root of f, then we have seen $f = (x-\alpha)(x+\alpha)(x-1/\alpha)(x+1/\alpha)$ over $\mathbb{Q}(\alpha)$, thus $\mathbb{Q}(\alpha)$ is the splitting field of f. As $\mathbb{Q}(\alpha)/\mathbb{Q}$ is a simple extension, we also know that $|Gal(f)| \leq 4$. On the other hand $\alpha \mapsto -\alpha$, $\alpha \mapsto 1/\alpha$ are obviously two commuting elements of $Gal(f)$ which generate a group isomorphic to $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$.

Exercises.

(1) Give an example of a simple algebraic extension K/k, such that $Gal(K/k) = \{1\}$, but $K \neq k$.

(2) Find the Galois group of $x^3 - 2x + 4$ over $\mathbb{Q}$.

9. The fundamental theorem of Galois theory

Now we will prove the famous fundamental theorem of Galois theory. It relates the intermediate fields L of a Galois extension K/k to the subgroups of the Galois group $Gal(K/k)$. The importance is that it relates questions in field theory to questions in group theory and in this way often makes it possible to solve them. The proof is not very long, because we already did a large portion of the work.

DEFINITION 9.1. Let K/k be a field extension. For any subgroup $H \subset Gal(K/k)$ the *fixed field* of H is

$$Fix(H) := \big\{a \in K \;\big|\; \varphi(a) = a \text{ for all } \varphi \in H\big\}.$$

REMARK 9.2.
(1) $Fix(H)$ is a subfield of K because, if $\varphi(a) = a$, $\varphi(b) = b$, then $\varphi(a \pm b) = \varphi(a) \pm \varphi(b)$, $\varphi(ab) = \varphi(a)\varphi(b)$, $\varphi(1/a) = 1/\varphi(a)$.
(2) If K/k is a Galois extension, then $Fix(Gal(K/k)) = k$.
(3) If L is an intermediate field of K/k, then

$$Gal(K/L) = \big\{\varphi \in Gal(K/k) \;\big|\; \varphi|_L = id\big\}$$

is a subgroup of $Gal(K/k)$.

THEOREM 9.3. *(Fundamental theorem of Galois theory) Let K/k be a Galois extension with Galois group G.*

(1) *The mappings $H \mapsto Fix(H)$, $L \mapsto Gal(K/L)$ are mutually inverse inclusion reversing bijections*

$$\big\{\textit{subgroups of } G\big\} \longleftrightarrow \big\{\textit{intermediate fields of } K/k\big\},$$

i.e. $Fix(Gal(L/k)) = L$, $\quad Gal(K/Fix(H)) = H$.
(2) *$[K : Fix(H)] = |H|$ and $[Fix(H) : k] = [G : H]$ for subgroups $H \subset G$.*
(3) *$[K : L] = |Gal(K/L)|$ and $[L : k] = [G : Gal(K/L)]$ for intermediate fields $k \subset L \subset K$.*

PROOF. The main part is (1). (2) and (3) are easy consequences. (1) Let L be an intermediate field of K/k. Then K/L is a Galois extension. (It is a splitting field of a polynomial over k and thus over L, thus K/L is a normal extension. Let f be the minimal polynomial of an element a of K over L and let g be the minimal polynomial of a over k. As K/k is separated g has no multiple roots in its splitting field. On the other hand f is a factor of g. Thus as g has no multiple roots in its splitting field. Thus K/L is separated.)

Therefore $Fix(Gal(K/L)) = L$ by the last remark. Let H be a subgroup of G. We put $F := Fix(H)$. We need to show $Gal(K/F) = H$. By the theorem of the primitive element $K = k(a)$ for some $a \in K$. Define

$$f := \prod_{h \in H} (x - h(a)) \in k(a)[x], \quad deg(f) = |H|.$$

All roots of f are distinct: $h_1(a) = h_2(a)$ implies $h_1 = h_2$, because $k(a)/k$ is a simple extension. We write $f = \sum_i b_i x^i$ with $b_i \in k(a)$. We want to see that $b_i \in F$, i.e. $f \in F[x]$. For $l \in H$ we put $l_*(f) = \sum_i l(b_i)x^i = \prod_{h\in H}(x - l(h(a)))$. Then obviously $l_*f = \prod_{h\in H}(x - h(a)) = f$. Thus we see that $b_i \in F[x]$ for all i, i.e. $f \in F[x]$. We have $f(a) = 0$, therefore the minimal polynomial g of a over F divides f. Since $k(a) = F(a)$, we see

$$|Gal(K/F)| = [K : F] = deg(g) \leq deg(f) = |H|.$$

But obviously H is a subgroup of $Gal(K/F)$, so $H = Gal(K/F)$.

(2) Let again $F = Fix(H)$. Since K is Galois over F we have $[K : F] = |Gal(K/F)| = |H|$. The second statement follows because $[K : F][F : k] = [K : k] = |G|$.

(3) The first statement is obvious because K/L is Galois. The second statement follows because $[K : L][L : k] = [K : k] = |G|$. □

Let K/k be a Galois extensions. The intermediate fields L/k correspond to the subgroups $Gal(K/L)$ of $Gal(K/k)$. Assume L/k is normal, does $Gal(K/L)$ have a special property?

With some thought we can guess what will happen: As L/k is normal we have the Galois group $Gal(L/k)$. We should be able to express it in terms of $Gal(K/k)$ its subgroup $Gal(L/k)$. The most natural choice is $Gal(L/k) = Gal(K/k)/Gal(K/L)$. As we know that $Gal(L/k)$ is a group, this would mean that $Gal(K/L)$ is a normal subgroup of $Gal(K/k)$.

It turns out that all this is true: L/k is normal if and only if $Gal(K/L)$ is a normal subgroup of $Gal(K/k)$ and in this case $Gal(L/k) = Gal(K/k)/Gal(K/L)$. This is the second part of the fundamental theorem of Galois theory.

NOTATION 9.4. Let L be an intermediate field of K/k, and let $\alpha \in G$. We write $\alpha(L) := \{\alpha(a) \mid a \in L\}$. This is clearly an intermediate field of K/k.

LEMMA 9.5. *Let L be an intermediate field of K/k and $\alpha \in G$. Then $Gal(K/\alpha(L)) = \alpha Gal(K/L)\alpha^{-1}$.*

PROOF. Let $\varphi \in G$. Then $\varphi \in Gal(K/\alpha(L))$ if and only if for all $a \in L$ we have $\varphi(\alpha(a)) = \alpha(a)$, i.e. $\alpha^{-1}\varphi\alpha \in Gal(K/L)$, i.e. $\varphi \in Gal(K/\alpha(L))$ if and only $\varphi \in \alpha Gal(K/L)\alpha^{-1}$. □

No we can show the relation between normal subgroups and normal extensions.

THEOREM 9.6. *Let K/k be a Galois extension and let L be an intermediate field. The following are equivalent:*

1. *L/k is normal,*
2. *$\alpha(L) = L$ for all $\alpha \in Gal(K/k)$*
3. *$Gal(K/L)$ is a normal subgroup of $Gal(K/k)$.*

PROOF. "(1)$\Longrightarrow$(2)" Let $a \in L$ and let f be the minimal polynomial of a over k. Since L is normal over k, all roots of f lie in L. Let $\alpha \in Gal(K/k)$, then $\alpha(a)$ is also a root of f because $f(\alpha(a)) = \alpha(f(a)) = 0$. Therefore $\alpha(a) \in L$ and $\alpha(L) \subset L$. The same argument shows $\alpha^{-1}(L) \subset L$, i.e. $L \subset \alpha(L)$.

"(2)$\Longrightarrow$(3)" Let $\alpha \in Gal(K/k)$. Then

$$\alpha Gal(K/L)\alpha^{-1} = Gal(K/\alpha(L)) = Gal(K/L).$$

Thus $Gal(K/L)$ is a normal subgroup of $Gal(K/k)$.

"(3)$\Longrightarrow$(1)" We have $Gal(K/L)$ is a normal subgroup of $Gal(K/k)$. Let $\alpha \in Gal(K/k)$. Then

$$Gal(K/L) = \alpha Gal(K/L)\alpha^{-1} = Gal(K/\alpha(L)).$$

By the fundamental theorem of Galois theory we get $L = \alpha(L)$. Let $f \in k[x]$ be irreducible with a root a in L, and let b be another root of f in K. Because both a and b have f as a minimal polynomial over k, we get that there is an $\alpha \in Gal(K/k)$ with $\alpha(a) = b$. Therefore $b \in \alpha(L) = L$. Thus L is normal over k. □

COROLLARY 9.7. *Let K/k be a Galois extension. Let L/k be an intermediate field which is Galois over k. Then*

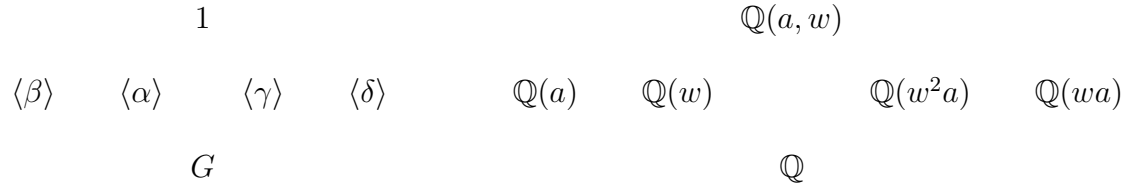
$$Gal(L/k) \simeq Gal(K/k)/Gal(K/L).$$

PROOF. Let $\alpha \in Gal(K/k)$. Then $\alpha(L) = L$, thus $\alpha|_L$ is an element of $Gal(L/k)$. The restriction map $Gal(K/k) \to Gal(L/k); \alpha \mapsto \alpha|_S$ is a homomorphism of groups with kernel $Gal(K/L)$. Thus by the first isomorphism theorem $Gal(K/k)/Gal(K/L)$ is isomorphic to a subgroup of $Gal(L/k)$. On the other hand

$$|Gal(K/k)/Gal(K/L)| = [K : k]/[K : L] = [L : k] = |Gal(L/k)|$$

because L/k is a Galois extension. Therefore $Gal(L/k) = Gal(K/k)/Gal(K/L)$. □

EXAMPLE 9.8. (1) Let $\mathbb{Q}(a, w)$ with $a = 2^{1/3}$, $w = e^{2\pi i/3}$ be the splitting field of $x^3 - 2 \in \mathbb{Q}[x]$. We have seen that $G := Gal(\mathbb{Q}(a, w)/\mathbb{Q}) = \mathcal{S}_3$. G has precisely 6 subgroups. Below we have on the left the diagram of subgroups H of G (a line indicates the upper group is a subgroup of the lower one) and on the right the corresponding diagram of intermediate fields $Fix(H)$ (a line indicates that the upper field is an extension of the lower one).



Here $\alpha(a) = wa$, $\alpha(w) = w$, $\beta(a) = a$, $\beta(w) = w^2$, $\gamma(a) = wa$, $\gamma(w) = w^2$, $\delta(a) = w^2a$, $\gamma(w) = w^2$. The normal subgroups of G are 1, $\langle\alpha\rangle$ and G corresponding to the normal extensions K, $\mathbb{Q}(w)$ and $\mathbb{Q}$.

(2) Let $f = x^4 + 1 \in \mathbb{Q}[x]$. Let α be a root of f, we know that over $\mathbb{Q}(\alpha)$ $f = (x-\alpha)(x+\alpha)(x-1/\alpha)(x-1/\alpha)$, and that $G := Gal(f) = \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, with the four elements 1, $\mu : \alpha \mapsto -\alpha$, $\delta : \alpha \mapsto 1/\alpha$, $\delta\mu : \alpha \mapsto -1/\alpha$. Thus we get the diagrams

$$\begin{array}{ccc} 1 & & \mathbb{Q}(\alpha) \\ \langle\mu\rangle \quad \langle\delta\rangle \quad \langle\mu\delta\rangle & & \mathbb{Q}(\alpha^2) \quad \mathbb{Q}(\alpha+1/\alpha) \quad \mathbb{Q}(\alpha-1/\alpha) \\ G & & \mathbb{Q} \end{array}$$

EXAMPLE 9.9. Let $f \in \mathbb{Q}[x]$ and let $K \subset \mathbb{C}$ be its splitting field. Let $\iota : \mathbb{C} \to \mathbb{C}; a + ib \mapsto \overline{a + i\overline{b}} := a - ib$ be complex conjugation. Denote by the same letter the restriction of ι to K. We claim that ι maps K to itself: As $\iota_* f = f$ it follows that if α is a root of f in K, then $\iota(\alpha)$ is a root of f. Thus $\iota(\alpha) \in K$. As K/k is generated by the roots of f, it follows that $\iota(K) \subset K$. Therefore $\iota|_K : K \to K$ is a k-isomorphism, as a restriction of a k-isomorphism. Thus we find that complex conjugation ι is an element of $Gal(K/k)$. Note that $K \subset \mathbb{R}$, if and only if $\iota = id_K$.

Let α be a zero of f in $K \setminus \mathbb{R}$. Then $\overline{\alpha} = \iota(\alpha) \neq \alpha$ is also a zero of f in $K \setminus \mathbb{R}$. Furthermore $\iota(\iota(\alpha)) = \alpha$. Thus we see that the nonreal zeros of f come in pairs α, $\overline{\alpha}$, in particular the number of zeros of f in $K \setminus \mathbb{R}$ is even.

Exercises.

(1) Determine the Galois group of $x^4 - 5$ over $\mathbb{Q}$. Determine the diagram of all subgroups and the diagram of all intermedate fields like in example 10.8.

(2) (a) Let $a \in \mathbb{Z}$. Prove that $x^3 + ax - 1$ is irreducible in $\mathbb{Q}[x]$ if and only if $a \neq 0$ and $a \neq -2$.
(b) Let s be a zero of $x^3 - 3x - 1$ in $\mathbb{C}$. Show that $-(1+s)^{-1}$ is also a zero of $x^3 - 3x - 1$.
(c) Let s as in part (b). Show that $\mathbb{Q}(s)/\mathbb{Q}$ is a Galois extension with Galois group isomorphic to $\mathbb{Z}/3\mathbb{Z}$.

(3) Determine the minimal polynomial for $i + \sqrt{2}$ over $\mathbb{Q}$.

(4) Determine the intermediate fields between $\mathbb{Q}$ and $Q(\sqrt{2}, \sqrt{3})$.

(5) Let $\sqrt[4]{2}$ be the positive real fourth root of 2. Factor the polynomial $x^4 - 2$ into irreducible factors over each of the fields $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(\sqrt{2}, i)$, $\mathbb{Q}(\sqrt[4]{2})$, $\mathbb{Q}(\sqrt[4]{2}, i)$.

(6) Let k be a field, let $K := k(\alpha)$ be a field extension of k with $\alpha^2 = a \in k$. Determine all elements of K whose squares are in k.

(7) Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$. Determine $[K : \mathbb{Q}]$, prove that $K/\mathbb{Q}$ is a Galois extension and determine its Galois group.

(8) Prove or disprove: Let $f \in \mathbb{Q}[x]$ be an irreducible polynomial of degree 3, with one real 0. The other roots are a pair $\beta, \overline{\beta}$ of complex conjugates, so that $L = \mathbb{Q}(\beta)$ has an automorphism, which exchanges β and $\overline{\beta}$.

(9) Let K/k be a Galois extension. Let $f \in k[x]$ be an irreducible polynomial. Show: All irreducible factors of f in $K[x]$ have the same degree.
(10) Let $f := x^4 + x^3 + x^2 + x + 1$, $\omega = e^{2\pi i/5}$. Put $K := \mathbb{Q}(\omega)$.
  (a) Using the Eisenstein criterion show that f is irreducible over $\mathbb{Q}$ and thus $[K : \mathbb{Q}] = 4$.
  (b) Show that $K/\mathbb{Q}$ is a Galois extension.
  (c) Show that an element of $Gal(K/\mathbb{Q})$ is determined by its value on ω.
  (d) Show that $Gal(K/\mathbb{Q})$ is a cyclic group of 4 elements
  (e) Determine all the intermediate fields of $K/\mathbb{Q}$.
(11) Let $f = x^n - 1 \in \mathbb{Q}[x]$. Show that the Galois group of f over $\mathbb{Q}$ is abelian.
(12)
  (a) Prove that the Galois group of $x^3 - 2$ over $\mathbb{Q}$ is $\mathcal{S}_3$ the symmetric group in 3 letters.
  (b) Find the splitting field of $x^3 - 2$ over $\mathbb{Q}$.
  (c) Write the diagram of subgroups of the Galois group and the correponding diagram of intermediate fields of the splitting field of $x^3 - 2$ over $\mathbb{Q}$.
(13) Determine the Galois group of the following polynomials over $\mathbb{Q}$.
  (a) $x^4 - 10x^2 + 5$.
  (b) $x^4 - x^2 - 6$.
(14) Let p be a prime number, $\omega = e^{2\pi i/5}$. Determine the Galois group of $\mathbb{Q}(\sqrt[5]{p}, \omega)$ over $\mathbb{Q}$.
(15) What are the degrees of the following fields over $\mathbb{Q}$?
  (a) $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.
  (b) $\mathbb{Q}(\sqrt{2}, \sqrt{-2})$.
  (c) $\mathbb{Q}(\sqrt[3]{2}, \sqrt{3})$.
  (d) $\mathbb{Q}(\sqrt[3]{2} - \sqrt{3})$.
(16) Which of the extensions of $\mathbb{Q}$ in the previous exercise are Galois extensions?
(17) Determine the Galois group of the splitting field over $\mathbb{Q}$ of $(x^3 - 2)(x^3 + 3)$.
(18) Let α be a complex root of the polynomial $x^3 + x + 1$ over $\mathbb{Q}$ and let K be the splitting field of this polynomial over $\mathbb{Q}$.
  (a) Is $\sqrt{-3}$ in $\mathbb{Q}(\alpha)$? Is it is K?
  (b) Prove that $\mathbb{Q}(\alpha)$ has no automorphism over $\mathbb{Q}$ except the identity.
(19) Let $K = \mathbb{Q}(\alpha)$, where α is a root of $x^3 + 2x + 1$ and let $g = x^3 + x + 1$. Does g have a root in K?
(20) Let $f \in k[x]$ be a polynomial of degree n, and let K be a splitting field for f over k. Prove that $[K : k]$ divides $n!$.
(21) Let G be a finite group. Prove that there exists a field k and a Galois extension K of k whose Galois group is G.
(22) Let $K/L/k$ be fields. Prove or disprove:
  (a) If K/k is Galois, then K/L is Galois.
  (b) If K/k is Galois, then L/k is Galois.

(c) If L/k and K/L are Galois, then K/k is Galois.

(23) Let K/k be a Galois extension whose Galois group is the symmmetric group $\mathcal{S}_3$. Is it true that K is the splitting field of an irreducible cubic polynomial over k?

(24) Let K/k be a Galois extension with Galois group G. Prove that there exists an element $\beta \in K$ whose stabilizer is H.

(25) Let K be a subfield of $\mathbb{C}$ which is a Galois extension of $\mathbb{Q}$. Prove of disprove: Complex conjugation carries K to itself, and thus it defines an automorphism of K.

10. Quadratic, biquadratic and Cubic polynomials

We illustrate the Main Theorem of Galois theory by analyzing the simplest classes of Galois extensions of a field k of characteristic 0, the splitting fields of quadratic, biquadratic and cubic polynomials.

Quadratic extensions.

The case of quadratic polynomials is quite obvious. Let $f = x^2 + px + q \in k[x]$ be an irreducible polynomial. We can write $f = (x + \frac{p}{2})^2 - (\frac{p^2}{4} - q)$. Thus the roots of f are $-\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$, $-\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}$. Put $\alpha = \sqrt{\frac{p^2}{4} - q}$. Then $k(\alpha)/k$ is a Galois extension, $[k(\alpha) : k] = 2$ and the Galois group is $\mathbb{Z}/2\mathbb{Z}$ generated be $\alpha \mapsto -\alpha$. Obviously there is no intermediate field between $k(\alpha)$ and k.

Biquadratic extensions.

Biquadratic extensions are not much for difficult. Let α be the zero of an irreducible monic polynomial $f \in k[x]$ and let β be the zero of an irreducible monic polynomial $g \in k[x]$, which is also irreducible over $k(\alpha)$. Then $[k(\alpha, \beta) : k(\alpha)] = 2 = [k(\alpha) : k]$. Note that this implies that also $[k(\alpha, \beta) : k(\beta)] = 2$. We see that $k(\alpha, \beta)$ is the splitting field of fg over k, thus $k(\alpha, \beta)/k$ is a Galois extension. Let α' be the other zero of f in $k(\alpha)$ and β' be the other zero of $g \in k(\beta)$. Put $a := \alpha - \alpha'$, $b := \beta - \beta'$. Then $k(a) = k(\alpha)$, $k(b) = k(\beta)$ and $k(a, b) = k(\alpha, \beta)$.

Because $k(\alpha, \beta)/k(\beta)$ is a simple extension, there is a unique automorphism φ of $k(\alpha, \beta)$ over $k(\beta)$ sending α to α', i.e. a to $-a$. Obviously $\varphi^2 = id$. In the same way there is a is a unique automorphism ψ of $k(a, b)$ over $k(a)$ sending b to $-b$ and $\psi^2 = id$. It is clear that $\varphi\psi = \psi\varphi$. Thus the Galois group of $k(\alpha, \beta)/k$ is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. Thus there are 3 proper subgroups: $H_1 = \{1, \varphi\}$, $H_2 = \{1, \psi\}$ and $H_3 = \{1, \psi\varphi\}$, and we see immediately that the corresponding fixed fields are $L_1 = k(a)$, $L_2 = k(b)$, $L_3 = k(ab)$.

Cubic extensions.

Finally we want to study the splitting fields of cubic equations. The explicit formulas for the roots of cubic equations in terms of cube roots and square roots were found in the sixteen century by Cardano and Tartaglia. We start by giving their

solution. Let $f : Y^3 + a_2Y^2 + a_1Y + a_0 \in k[Y]$. By the substitution $Y = x - a_2/3$ we can assume that the quadratic term vanishes, i.e.

$$f = x^3 + px + q, \quad p, q \in k.$$

The trick is to substitute $x = U - V$. This does not look like a good idea: the equation only becomes more complicated, however we shall see in a moment why it helps. We get

$$f(U - V) = U^3 - V^3 - (3UV - p)(U - V) + q.$$

The point of this substitution is that we can split the equation appart. Obviously a sufficient condition for $f(U - V) = 0$ is that both equations

$$U^3 - V^3 + q = 0,$$
$$3UV - p = 0$$

hold. The second equation gives $V = p/(3U)$, which we substitute into the first equation. This gives

$$3^3U^6 - p^3 + 3^3U^3q = 0.$$

Now something very nice has happened: This equation is quadratic in U^3, so we can solve it: put $Y := U^3$, then the equation is $3^3Y^2 + 3^3qY - p^3 = 0$, which we can solve by the quadratic case: one solution is

$$Y = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}.$$

Thus one solution for U is just $U = \sqrt[3]{Y}$ and $V = \sqrt[3]{U^3 + q}$. Putting this together we get the famous formula of Cardano:

One solution is

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} - \sqrt[3]{\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

Putting this into the equation one can check that this is indeed a solution.

Now we study the Galois theory for the irreducible cubic $f = x^3 + px + q$ over k. Let K be a splitting field of f over k. Thus we can write $f = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$, which gives

$$\alpha_1 + \alpha_2 + \alpha_3 = 0,$$
$$\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_1\alpha_3 = p,$$
$$\alpha_1\alpha_2\alpha_3 = -q.$$

The first equation gives that $\alpha_3 \in k(\alpha_1, \alpha_2)$. Thus we have a chain of fields

$$k \subset L := k(\alpha_1) \subset k(\alpha_1, \alpha_2) = K.$$

There are two possibilities, either $L = K$ or $L \subsetneq K$.

We want to find out when these two cases happen. For this we have to check how f factors in $L[x]$. As f is irreducible over k we see that $[L : k] = 3$. In $L[x]$ we can

write $f = (x - \alpha_1)g$ with $g = (x - \alpha_1)(x - \alpha_2) \in K[x]$. Thus if g is not irreducible in $L[x]$, then $L = K$ and $[K : k] = 3$, otherwise $[K : L] = 2$ and thus $[K : k] = 6$.

(1) If $K = L$, then K/k is a simple algebraic extensions, thus $Gal(K/k)$ is a subgroup of $\mathcal{S}_3$ which acts simply transitively on $\alpha_1, \alpha_2, \alpha_3$. The only such subgroup is the set of cyclic permutations of $\alpha_1, \alpha_2, \alpha_3$. As this has no nontrivial subgroup, there are no intermediate fields between K and k.
(2) If $K \neq L$, then $Gal(K/k)$ is the symmetric group $\mathcal{S}_3$ acting as group of permutations of $\alpha_1, \alpha_2, \alpha_3$.

We can decide which of the two happens in terms of the discriminant. We will introduce this for polynomials of arbitrary degree.

DEFINITION 10.1. Let $f \in k[x]$ be an irreducible polynomial of degree n. Let K be its splitting field over k, let $\alpha_1, \dots, \alpha_n$ be its roots in K.

$$D := \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)^2.$$

We put $\delta := \sqrt{D} = \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)$. Note that D is invariant under all permutations of $\alpha_1, \dots, \alpha_n$. Thus D is fixed by $Gal(K/k)$, thus $D \in k$. In case $f = x^3 + px + q$ we can compute directly that $D = -4p^3 - 27q^2$.

The discriminant now tells us whether $Gal(K/k)$ only consists of even permutations of the α_i.

PROPOSITION 10.2. *$Gal(K/k) \subset A_n$ if and only if D is a square in k (or equivalently $\delta \in k$).*

PROOF. Let $\sigma \in \mathcal{S}_n$. Then $\sigma(\delta)$ is $(-1)^k \delta$, where k is the number of pairs $i < j$ with $\sigma(i) > \sigma(j)$, i.e. $\sigma(\delta) = sign(\sigma)\delta$, where $sign(\sigma)$ is the sign of the permutation. Thus δ is fixed by all $\sigma \in Gal(K/k)$ if and only if $Gal(K/k)$ is a subgroup of the permutations of sign 1, i.e. A_n. □

Now we go back to the case that $f = x^3 + px + q$ is an irreducible cubic with $Gal(f) = \mathcal{S}_3$. We analyse the subgroups and intermediate fields. $\mathcal{S}_3$ has 3 conjugate subgroups of order 2 (generated by one transposition) and one subgroup A_3 (the set of cyclic permutations) of order 3. The intermediate fields $k(\alpha_1), k(\alpha_2), k(\alpha_3)$ are the fixed fields of the transpositions $\alpha_2 \mapsto \alpha_3$, $\alpha_1 \mapsto \alpha_3$, $\alpha_1 \mapsto \alpha_2$, and $k(\delta)$ is the fixed field of the set of cyclic permutations.

Exercises.

(1) Prove that the discriminant of a cubic $f \in \mathbb{R}[x]$ is positive if all the roots are real and negative if not.
(2) Determine the Galois groups of the following polynomials over $\mathbb{Q}$.
   (a) $x^3 + 27x - 4$.
   (b) $x^3 + x + 1$.
   (c) $x^3 + x^2 - 2x + 1$.

(d) $x^3 + x^2 - 2x - 1$.

(3) Let f be an irreducible cubic polynomial over k and let δ be the square root of the discriminant. Show that f remains irreducible over $k(\delta)$.

(4) Let $f \in \mathbb{Q}[x]$ be an irreducible cubic polynomials which has precisely one real root, and let K be the splitting field of f over $\mathbb{Q}$. Show that $[K : \mathbb{Q}] = 6$.

(5) Prove that the discriminant of the cubic $x^3 + px + q$ is $-4p^3 - 27q^2$.

11. Solvability by radicals

In this whole section let k be a field of characteristic 0.

The main problem that motivated the development of Galois theory was the question of solvability of polynomials by radicals. Given a polynomial

$$f = x^n + \sum_{i=0}^{n-1} a_i x^i \in \mathbb{Q}[x],$$

one wants to find a formula for the roots of f in $\mathbb{C}$, in terms of the a_i. Here the operations in the formula are just $+$, $-$, $\cdot$, $/$ and $\sqrt[r]{\ }$. This problem has been studied over the centuries by many people. For $n = 2$ we have the classical formula for the roots of $x^2 + ax + b$ by $-\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b}$. For $n = 3$ and $n = 4$ the formulas were found in the 16^{th} century. For $n \geq 5$ the problem remained open, until Galois theory was invented. We start by giving a precise mathematical formulation of the problem.

Definition 11.1. A field extension K/k is called a *radical extension* if there is a chain

$$K = L_m \supset L_{m-1} \supset \ldots \supset L_0 = k$$

of intermediate fields, so that $L_{i+1} = L_i(b_i)$ for all i, where b_i is a root of a polynomial $x^{m_i} - a_i$ with $a_i \in L_i$ (thus b_i is the m_i^{th} root of an element of L_i).

We want to express the roots of $f \in k[x]$ in terms of the usual operations of a field and by taking i^{th} roots. Thus we say that f is solvable by radicals if it splits into linear factors over a radical extension of k.

Definition 11.2. A polynomial $f \in k[x]$ is *solvable by radicals* if there is a radical extension K/k such that K/k is a Galois extension and f splits over K into linear factors.

Remark 11.3. The condition that K/k is a Galois extension is not natural, we make it because it simplifies our proofs. One can show that, if K/k is radical extension, then there is a radical extension F/k with $K \subset F$, such that F/k is a Galois extension. This is an exercise to this section. Thus it follows that in the definition above we can drop the condition that K/k is Galois.

This question whether f is solvable by radicals is closely related to the structure of the Galois group of f. Therefore we will need to develop a little bit more of group theory.

DEFINITION 11.4. A group G is called *solvable* if there is a chain

$$G = G_0 \supset G_1 \supset \ldots \supset G_n = \{1\}$$

of subgroups of G such that G_{i+1} is a normal subgroup in G_i for all i and the factor group G_i/G_{i+1} is abelian. (in particular abelian groups are solvable).

Solvable groups play an important role in the theory of groups. One could say that being solvable is a generalization of abelian. The name *solvable* comes from its relation to solvability by radicals.

EXAMPLE 11.5. One can show that $\mathcal{S}_n$ is solvable for $n \leq 4$, but not solvable for $n \geq 5$.

We need a criterion for the solvablity of a group. For this we introduce the commutator subgroup of a group.

DEFINITION 11.6. Let G be a group, and let $a, b \in G$. The *commutator* of a, b is $[a, b] = a^{-1}b^{-1}ab$. Note that $[a, b] = 0$ if and only if $ab = ba$. The commutator subgroup G' of G is the subgroup of G generated by $\big\{[a, b] \;\big|\; a, b \in G\big\}$.

It is clear that G is abelian if and only if $G' = \{1\}$.

The following Lemma will help us find a relation between solvability of a group and commutator subgroups.

LEMMA 11.7. (1) *G' is a normal subgroup of G.*
(2) *G/G' is abelian*
(3) *If H is a subgroup of G with G/H abelian, then $G' \subset H$.*

PROOF. (1) Note that

$$g^{-1}[a, b]g = g^{-1}a^{-1}gg^{-1}b^{-1}gg^{-1}agg^{-1}bg = [g^{-1}ag, g^{-1}bg].$$

Note that $[a, b]^{-1} = [b, a]$. Thus the elements of G' are just products of commutators of elements of G. Thus let $h = [a_1, b_1][a_2, b_2] \ldots [a_n, b_n] \in G'$. Let $g \in G$. Then

$$g^{-1}hg = g^{-1}[a_1, b_1]g \ldots g^{-1}[a_n, b_n]g = [g^{-1}a_1g, g^{-1}b_1g] \ldots [g^{-1}a_ng, g^{-1}b_ng] \in G'.$$

Thus G' is a normal subgroup of G. (2) Let $a, b \in G$. Then $aG' \cdot bG' = abG' = ba[a, b]G' = baG' = bG' \cdot aG'$. (3) Assume $aH \cdot bH = bH \cdot aH$ for all $a, b \in G$. Then $baH = abH = ba[a, b]H$. Thus $[a, b] \in H$. Thus $[a, b] \in H$ for all $a, b \in G$. As G' is the subgroup generated by all $[a, b]$, it follows that $G' \subset H$. □

We can iterate taking the commutor:

DEFINITION 11.8. Let $G^{(0)} = G$, $G^{(1)} = G'$ and inductively $G^{(n+1)} = (G^{(n)})'$. Note that $G^{(n+1)}$ is a normal subgroup in $G^{(n)}$ and $G^{(n)}/G^{(n+1)}$ is abelian for all n.

This gives a criterion for a group to be solvable:

LEMMA 11.9. *G is solvable if and only if $G^{(n)} = \{1\}$ for some n.*

PROOF. Assume $G^{(n)} = \{1\}$. Then we have a chain of subgroups

$$G = G^{(0)} \supset G^{(1)} \supset \ldots \supset G^{(n)} = \{1\}$$

and by the previous lemma we have $G^{(i+1)}$ is a normal subgroup in $G^{(i)}$ and $G^{(i)}/G^{(i+1)}$ is abelian. Thus G is solvable.

Conversely assume G is solvable. Then there is a chain

$$G = H_0 \supset H_1 \supset \ldots \supset H_n = \{1\}$$

of subgroups of G with H_{i+1} a normal subgroup of H_i and H_i/H_{i+1} abelian. But then by the previous lemma $H_i' \supset H_{i+1}$. Thus $H_1 \supset H_0' = G'$, $H_2 \supset H_1' \supset G'' = G^{(2)}$ and inductively we get $\{1\} = H_n \supset G^{(n)}$, i.e. $G^{(n)} = \{1\}$. □

COROLLARY 11.10. *Let G be a solvable group and $\varphi : G \to H$ a surjective group homomorphism. Then H is solvable.*

PROOF. Obviously $\varphi([a,b]) = [\varphi(a), \varphi(b)]$. It follows that $\varphi(G') = H'$, and inductively $\varphi(G^{(i)}) = H^{(i)}$ for all i. Since $G^{(n)} = \{1\}$ for some n it follows that $H^{(n)} = \varphi(\{1\}) = \{1\}$. □

Now we want to study the solvability of a polynomial $f \in k[x]$ by radicals. As a first step we have to deal with roots of unity.

DEFINITION 11.11. Let $n \in \mathbb{Z}_{>0}$. Let K be a field. An element $\zeta \in K$ is called an n-th root of unity if $\zeta^n = 1$. It is called a primitive n-th root of unity if $\zeta^n = 1$ and $\zeta^m \neq 1$ for all $0 < m < n$. (If $K = \mathbb{C}$ a primitive n-th root of unity is for instance $e^{2\pi i/n}$).

LEMMA 11.12. *Let $n \in \mathbb{Z}_{>0}$. Let k be a field. Let ζ (in some extension of k) be a primitive n-th root of unity. Then $k(\zeta)/k$ is a Galois extension and $Gal(k(\zeta)/k)$ is abelian.*

PROOF. Over $k(\zeta)$ the polynomial $X^n - 1$ splits into linear factors: the ζ^r, $r = 0, \ldots n-1$ are all distinct zeros of $X^n - 1$ in $k(\zeta)$. If $\zeta^i = \zeta^j$, with $0 \leq i < j < n$, then $\zeta^{j-i} = 1$, a contradiction to ζ being primitive. It follows that $k(\zeta)$ is the splitting field of $X^n - 1$ over k, thus $k(\zeta)/k$ is a Galois extension.

Let $G = Gal(k(\zeta)/k)$. If $\sigma \in G$, then $\sigma(\zeta)$ must be a root of $X^n - 1$, thus $\sigma(\zeta) = \zeta^s$ for some s. We know that G is a subgroup of the permutations of $\{\zeta^0, \ldots, \zeta^{n-1}\}$. It follows that two elements $\sigma, \tau \in G$ are equal if $\sigma(\zeta) = \tau(\zeta)$.

Let $\sigma, \tau \in G$. Then $\sigma(\zeta) = \zeta^s$, $\tau(\zeta) = \zeta^t$ for some s, t, and

$$\sigma\tau(\zeta) = \sigma(\zeta^t) = \sigma(\zeta)^t = \zeta^{st} = \tau\sigma(\zeta).$$

Thus $\sigma\tau = \tau\sigma$. □

THEOREM 11.13. *Let n be a positive integer. Let k be a field that contains a primitive n-th root of unity. Let $a \in k \setminus \{0\}$. Let K be the splitting field of $x^n - a$ over k.*

(1) *$K = k(\alpha)$ where α is any root of $x^n - a$ in K.*
(2) *$Gal(K/k)$ is abelian.*

PROOF. (1) Let α be a root of $x^n - a$ in K, then $\alpha, \zeta\alpha, \dots, \zeta^{n-1}\alpha$ are all the roots of $x^n - a$: Obviously they are roots of $x^n - a$. Thus it is enough that they are distinct. If $\zeta^i\alpha = \zeta^j\alpha$ with $0 \le i < j < n$, then $(\zeta^i - \zeta^j)\alpha = 0$, thus $\zeta^i = \zeta^j$ i.e. $\zeta^{j-i} = 1$, which is impossible. Thus we see that all roots of $x^n - a$ are in $k(\alpha)$, i.e. $K = k(\alpha)$.

(2) Let $\sigma, \tau \in Gal(K/k)$. Then $\sigma(\alpha)$ and $\tau(\alpha)$ are roots of $x^n - a$, thus $\sigma(\alpha) = \zeta^i\alpha$, $\tau(\alpha) = \zeta^j\alpha$ for some i, j. Thus

$$\sigma\tau(\alpha) = \sigma(\zeta^j\alpha) = \zeta^j\sigma(\alpha) = \zeta^j\zeta^i\alpha = \zeta^{i+j}\alpha,$$

similarly $\tau\sigma(\alpha) = \zeta^{i+j}\alpha$. Thus $\sigma\tau$ and $\tau\sigma$ agree on α and on k, and therefore on $K = k(\alpha)$. Thus $\sigma\tau = \tau\sigma$; thus the Galois group is abelian. □

This theorem tells us, that if the field k contains enough roots of unity, then for each $a \in k$ the Galois group of $x^n - a$ over k is abelian.

THEOREM 11.14. *(Galois) Let k be a field of characteristic 0. Let $f \in k[x]$ and assume f is solvable by radicals over k.*

Then the Galois group $Gal(f)$ over k is solvable.

REMARK 11.15. Conversely one can show that f is solvable by radicals over k if $Gal(f)$ is solvable.

PROOF. Let K be the splitting field of f over K. Then $Gal(f) = Gal(K/k)$ and we have to show that $Gal(K/k)$ is solvable. Since f is solvable by radicals, there exists a sequence

$$k = k_1 \subset k_2 = k(\alpha_2) \subset k_3 = k_2(\alpha_3) \subset \ldots \subset k_n = k_{n-1}(\alpha_n),$$

such that for each i there exists an $r_i > 0$ with $a_i := \alpha_i^{r_i} \in k_{i-1}$ and f splits over k_n into linear factors, i.e. $K \subset k_n$. Furthermore k_n/k is a Galois extension.

Let m be the greatest common multiple of the r_i. Let ζ be a primitive m-th root of unity, and let $F := k(\zeta)$. Then ζ^{m/r_i} is a primitive r_i-th root of unity, thus F contains primitive r_i-th roots of unity for all i. We put $F_0 = k$, $F_1 = F$, $F_i = F_{i-1}(\alpha_i)$ for all i. Thus we get a chain

$$F = F_0 \subset F_1 \subset \ldots \subset F_n,$$

such that $Gal(F_i/F_{i-1})$ is abelian and $K \subset F_n$. k_n is a splitting field of some polynomial g over k and F_n is the splitting field of $(X^m - 1)g$. Thus F_n/k is a Galois extension.

Therefore F_n/F_i is also a Galois extension for all $i = 0, \dots, n$. F_1 and therefore also all F_{i-1} with $i > 1$ contain all r_i-th roots of unity. By the previous theorem $F_i = F_{i-1}(\alpha_i)$ is the splitting field of $x^{r_i} - a_i$ over F_{i-1} and thus F_i/F_{i-1} is a Galois extension. We also know F_1/F_0 is a Galois extension. By the second part of the fundamental Theorem of Galois theory we know that $Gal(F_n/F_i)$ is a normal subgroup of $Gal(F_n/F_{i-1})$, and $Gal(F_n/F_{i-1})/Gal(F_n/F_i) = Gal(F_i/F_{i-1})$, which is abelian. Therefore the chain

$$G(F_n/k) \supset G(F_n/F_1) \supset \ldots \supset G(F_n/F_{n-1})$$

shows that $G(F_n/k)$ is solvable! We have that K is a subfield of F_n, and K/k is a normal extension, because K is a splitting field of a polynomial over k. Thus again by the second part of the principal theorem of Galois theory we get $Gal(K/k)$ is isomorphic to $Gal(F_n/k)/Gal(F_n/K)$. Thus there is a surjective homomorphism $Gal(F_n/k) \to Gal(K/k)$. By a lemma we proved before, $Gal(f) = Gal(K/k)$ is also solvable. This proves the theorem. □

This result does not look very useful if we want to find out whether a given polynomial is solvable or not. How are we ever going to find out whether the Galois group is solvable or not? Now we want to give a simple criterion that for many polynomials shows that they are not solvable. In fact we will give a simple way to check via elementary calculus for many polynomials $f \in \mathbb{Q}[x]$ of prime degree p that they have Galois group $\mathcal{S}_p$. As $\mathcal{S}_p$ is not solvable for $p \geq 5$, this implies that f is not solvable.

THEOREM 11.16. *Let p be a prime number and $f \in \mathbb{Q}[x]$ an irreducible polynomial of degree p. Assume f has precisely $p-2$ roots over $\mathbb{R}$. Then $Gal(f) = \mathcal{S}_p$.*

PROOF. Let $\Sigma = \{a_1, \dots, a_p\}$ be the set of roots of f, with $a_3, \dots, a_p \in \mathbb{R}$. Write L for the splitting field of f. We know that $Gal(f)$ is a subgroup of $S(\Sigma)$, which we identify with $\mathcal{S}_p$. Furthermore $Gal(f)$ acts transitively on Σ. As f is irreducible of degree p, we have $[\mathbb{Q}(a_1) : \mathbb{Q}] = p$ and thus $|Gal(f)| = [L : \mathbb{Q}] = [L : \mathbb{Q}(a_1)][\mathbb{Q}(a_1) : \mathbb{Q}]$ is divisible by p. By Cauchy's Theorem there exists an element $\rho \in Gal(f)$ of order p. The only elements of $\mathcal{S}_p$ of order p are the cyclic permutations. Thus $Gal(f)$ contains all cyclic permutations ρ^k. Replace ρ by a suitable power, so that $\rho(a_1) = a_2$. By reordering the other roots we can assume that $\rho(a_{k-1}) = a_k$ for all k. For the complex conjugation $\tau \in Gal(f)$ we have $\tau(a_1) = a_2$, $\tau(a_2) = a_1$ and $\tau(a_k) = a_k$ for $k \neq 1, 2$. Thus for $2 \leq j \leq p$ we have that $\rho^{j-1}\tau\rho^{1-j} \in Gal(f)$, and

$$\rho^{j-1}\tau\rho^{1-j}(a_i) = \rho^{j-1}\tau a_{i+1-j} = \rho^{j-1}a_{i+1-j} = a_i,$$

if $i + 1 - j \neq 1, 2$, i.e. $i \notin \{j, j+1\}$. On the other hand

$$\rho^{j-1}\tau\rho^{1-j}(a_j) = \rho^{j-1}\tau a_1 = \rho^{j-1}a_2 = a_{j+1},$$
$$\rho^{j-1}\tau\rho^{1-j}(a_{j+1}) = \rho^{j-1}\tau a_2 = \rho^{j-1}a_1 = a_j.$$

Thus $Gal(f)$ contains for all j the transposition $(j, j+1)$. These elements obviously generate $\mathcal{S}_p$, thus $Gal(f) = \mathcal{S}_p$. □

EXAMPLE 11.17. By elementary calculus it is easy to check e.g. for many polynomials of degree 5 that they have 3 real roots. For example:

(1) $f = 2x^5 - 10x + 5$. By the Eisenstein criterion f is irreducible over $\mathbb{Q}$. One checks by elementary calculus that f has a minimum at $x = 1$, with $f(1) = -3 < 0$, a maximum at $x = -1$ with $f(-1) = 3 > 0$ and no other extremal values. Obviously $f(x)$ tends to $+\infty$ for $x \to +\infty$ and $f(x)$ tends to $-\infty$ for $x \to -\infty$. It follows by the intermediate value theorem that f has precisely 3 real roots, One somewhere in $(-\infty, -1)$, one somewhere in $(-1, 1)$ and one in $(1, \infty)$.
(2) $f = x^5 - 6x + 3$ is not solvable by radicals. (Check as an exercise that f is irreducible over $\mathbb{Q}$ and has 3 real roots).

EXERCISE 11.18. (1) Let k be a field of characteristic 0.

(a) Let $k(\alpha_1, \dots, \alpha_n)/k$ be a finite field extension. Let L/K be a field extension, such that L/k is a finite Galois extension (e.g. let L be the splitting field of the product of the minimal polynomials of the α_i). Assume K/k is not normal. We put

$$K' = k\big(g(\alpha_i) \;\big|\; i = 1, \dots, n,\ g \in Gal(L/k)\big).$$

Show that K'/k is a normal extension. (Hint: Note that K' being a normal extension is equivalent to $g(K') = K'$ for all $g \in Gal(L/k)$), and this is essentially obvious.)

(b) Now assume that each α_i satisfies an equation $X^{r_i} - a_i = 0$ with $a_i \in k(\alpha_1, \dots, \alpha_{i-1})$. Then we have for all $g \in Gal(L/k)$ that $g(\alpha_i)$ satisfies the equation $X^{r_i} - g(a_i) = 0$, with $g(a_i) \in k(g(\alpha_1), \dots, g(\alpha_{i-1}))$. Conclude that K'/k is a radical Galois extension.