Basis Algebraic Structures

Lecture Notes for the MICS

by Martin Schlichenmaier

put to LaTeXby Robert Földes

March 2011, corrected and updated version January 2019

Contents

Literature for Basic Algebraic Structure

It is not so easy to give good recommendations. The following is rather incomplete. As a general guide-line, let me recommend books on **Linear Algebra** as a first starting point. Normally there the basic definitions on groups, rings, fields, vector spaces can be found. For more detailed information one has to get information out of books on **Algebra**. But typically these books contain much more than the material of relevance for the course.

1. Fisher, Gerd: Lineare Algebra, vieweg studium, is not expensive, but unfortunately it is in German. Of special relevance are the pages 32-74, but also the basic definitions related to vector spaces (75-163).

2. van der Waerden: Algebra I, Springer (this book is in German, but there is also an English translation). This is a classics. Not too difficult to read.

3. Wüstholz, Gisbert: Algebra für Studierende der Mathematik, Physik und Informatik, vieweg studium (only in German).

4. Lang, Serge: Algebra, Springer (English), in some sense a standard book, but contains much more than is needed here.

5. Griphone, Joseph: Algèbre linéaire, Cépaduès-éditions, Toulouse (this book is in French).

6. Shoup, Victor: A computational introduction to number theory and algebra (also available online at http://shoup.net/ntb).

7. Lüttkebohmert, W.: Codierungstheorie, vieweg, 2003 (this book is in German).

8. Witt, Kurt-Ulrich: Algebraische Grundlagen der Informatik, vieweg (this book contains more or less everything we need for the lecture, but it is in German - and I spotted some small errors).

1 Notation and conventions

Without further mentioning, we will use following mathematical conventions.

Sets: sets are given by the objects which constitute their elements. They are either given by a list or by definitions, e.g.
$\mathbb{N}$: the set of all natural numbers (Convention: they are bigger than 0),
$\mathbb{N}_0$: the set of all integers ≥ 0.,
$\mathbb{Z}$: all integers,
$\mathbb{Q}$: rational numbers,
$\mathbb{R}$: real numbers,
$\mathbb{C}$: complex numbers.

If M is a set, then $a \in M$ says that a is an element of M; $a \notin M$ says that a is not an element of M.

$$\mathbb{R}^{\geq 0} := \{\alpha \in \mathbb{R} \mid \alpha \geq 0\}$$

(all real numbers α with $\alpha \geq 0$).

Logical constructions:

$$A \wedge B : \qquad \text{(A and B)},$$
$$A \vee B : \qquad \text{(A or B)}.$$

"$\forall x \in M$:" For all elements $x \in M$ the following is true.
"$\exists x \in M$:" There exists $x \in M$ such that the following is true

Operations for sets:

$$M \cup N := \{x \mid x \in M \text{ or } x \in N\} \qquad \text{(union)},$$

$$M \cap N := \{x \mid x \in M \text{ and } x \in N\} \qquad \text{(intersection)},$$

If $M \subset A$, then the compliment is given as $C_A(M) := \{x \in A \mid x \notin M\}$.

De Morgan rules

$$C_A(M) \cap C_A(N) = C_A(M \cup N),$$
$$C_A(M) \cup C_A(N) = C_A(M \cap N).$$

(There are more laws for sets: distributive law.)

Maps:
Given A and B two sets. A map $f : A \to B$ is by definition an assignment to every element $x \in A$ of a unique element $y \in B$. y is denoted by $y = f(x)$.

$$f : A \to B; \quad x \mapsto y = f(x).$$

Warning: two different x_1 and x_2 can map to the same y.

Composition of maps:
Let A, B and C be sets. $f : A \to B$, $g : B \to C$ can be composed to a map:

$$k := g \circ f : A \to C,$$

("g after f") by defining

$$(g \circ f)(x) = g(f(x)).$$

Important fact: the composition of maps is associative, i.e.

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

The image of a map $f : A \to B$ is defined as

$$\mathrm{Im}f := \{y \in B \mid \exists x \in A : y = f(x)\} \subset B.$$

A map $f : A \to B$ is called

1. surjective if and only if $\mathrm{Im}f = B$.
2. injective if and only if $f(x_1) = f(x_2) \to x_1 = x_2$.
3. bijective if and only if f is surjective and injective.

Every bijective map $f : A \to B$ admits a unique inverse maps

$$f^{-1} : B \to A.$$

It fulfills $f^{-1} \circ f = id_A$, $f \circ f^{-1} = id_B$. Here $id_A : A \to A$ is the identity map $x \mapsto x$. For $f^{-1}(y)$ there is a unique element $x \in A$ such that $y = f(x)$ and it is set $f^{-1}(y) = x$.

We will need in a later part of the course also the definition of a vector space over a field (like $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$,...). This includes the definition of a basis, of dimension, linear maps, etc. Please consult for this books on linear algebra.

2 Groups

2.1 Basic definitions and examples

Groups are the mathematical structure behind symmetries. The rotations in the plane around an origin constitute a group. But also numbers with their arithmetic operations often build a group.

Definition 2.1. Let G be a set and $* : G \times G \to G$ a map (sometimes called composition, sometimes product).

1. $*$ is called associative if and only if $\forall a, b, c \in G : (a * b) * c = a * (b * c)$.
2. $*$ is called commutative if and only if $\forall a, b \in G : a * b = b * a$.
3. $e \in G$ is called a unit of $*$ if and only if $\forall a \in G : e * a = a * e = a$.
4. If $*$ has a unit e, then a' is called an inverse element of a if and only if $a * a' = a' * a = e$.

Definition 2.2. Let $(G, *)$ be a set G with a product $*$. This pair is called a group if:

1. $*$ is associative.
2. $*$ has a unit e.
3. every element $g \in G$ has an inverse g'.

Note that the group law is not assumed to be commutative. A group which has a commutative group law is called *commutative group*.

Proposition 2.3. *Given $(G, *)$ a group. Then:*

1. *e is unique.*
2. *The inverse of an element is unique.*

Proof.
ad 1. Let e and e' be units. Then $e * e' = e'$, as e is a unit. But also: $e * e' = e$, as e' is a unit. Hence $e' = e$.

ad 2. Let us assume that g' and g'' are two inverse elements of g. Hence:

$$\begin{aligned} g'' &= g'' * e && \text{as } e \text{ is unit} \\ &= g'' * (g * g') && \text{as } g' \text{ is the inverse to } g \\ &= (g'' * g) * g' && \text{as the product is associative} \\ &= e * g' && \text{as } g'' \text{ is the inverse to } g \\ &= g'. && \text{as } e \text{ is unit} \end{aligned}$$

Hence: $g'' = g'$.

□

By unicity of the inverse of an element, one can speak of <u>the</u> inverse element of g. We denote it by g^{-1}.

Remark 1. on notation:
1. Depending on the situation, one sometimes also uses the word "neutral element" for the unit element.
2. To use the symbol $*$ for the "product" is completely arbitrary. One can also use symbols like $+$, $\cdot$, $\circ$, ...
3. Quite often if one uses $+$ for the product one also denotes the inverse element of g by $(-g)$.
4. A commutative group is sometimes also called abelian group, in honour of the great Norwegian mathematican Niels Abel.

Example 2.4. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$.
Here, $\mathbb{Z}$ the integers, $\mathbb{Q}$ the rational numbers, $\mathbb{R}$ the real numbers, $\mathbb{C}$ are the complex numbers and "+" is the usual addition of numbers. These groups are all commutative. The unit element is the number 0, and the inverse of g is $(-g)$.

Example 2.5. $(\mathrm{Rot}(2), \circ)$
The rotations in the real plane $\mathbb{R}^2$ around the origin $0 = (0, 0)$, and "$\circ$" is the "composition" of rotations. This means that if we denote the elements by R_φ, rotation by the angle φ, then $R_{\varphi_1} \circ R_{\varphi_2}$ means first rotate by φ_2 and then by φ_1. The result will be a rotation by the angle $\varphi_1 + \varphi_2$ around $0 = (0, 0)$. The unit element is R_0, the rotation by zero degree and $(R_\varphi)^{-1} = R_{-\varphi}$. As $R_{\varphi_1} \circ R_{\varphi_2} = R_{\varphi_2} \circ R_{\varphi_1}$, the group is commutative.

Example 2.6. $(\mathrm{Rot}(3), \circ)$
Rotations in the three dimensional space $\mathbb{R}^3$ around $0 = (0, 0, 0)$. This group is not

commutative.

Example 2.7. Let $\{1, 2, 3\}$ be the set consisting of the three numbers 1, 2 and 3. Consider the set S_3 of all bijective mappings $\sigma : \{1, 2, 3\} \to \{1, 2, 3\}$. Define as product $\circ$ on S_3 the composition of maps. Then $(S_3, \circ)$ is a group which is non-commutative. This can be extended to sets with n numbers $1, 2, \ldots, n$. In his way one obtains the groups $(S_n, \circ)$, $n \in \mathbb{N}$.

Exercise 2.8. Check in the Examples 2.4 and 2.5 that the listed sets are indeed groups. (Verify the conditions in 2.1.)

Remark 2. $(\mathbb{N}, +)$, the natural numbers with addition as map is not a group.
Why? But there is a weaker concept, the concept of a monoid.
$(\mathbb{N}, +)$ is a monoid. We can always add two natural numbers to obtain another natural number and the addition is associative, i.e. we have: $\forall n, m, k \in \mathbb{N} : (n + m) + k = n + (m + k)$.
If we consider $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, we obtain a monoid with unit.

Definition 2.9. Let $(G, *)$ be a group. A subset $H \subseteq G$ is called a subgroup of G if and only if

1. $H \neq \varnothing$.
2. $g_1, g_2 \in H \Rightarrow g_1 * g_2 \in H$.
3. $g \in H \Rightarrow g^{-1} \in H$.

Remark 3.

1. The conditions 2 and 3 in 2.9 can be combined to:
$g_1, g_2 \in H \Rightarrow g_1 * g_2^{-1} \in H$.

2. If e is the unit element in G and H a subgroup of G, then $e \in H$.

Proof. H non-empty $\Rightarrow \exists g \in H \Rightarrow g^{-1} \in H \Rightarrow e = g * g^{-1} \in H$. □

3. If H is a subgroup of $(G, *)$, then $(H, *_{|H})$ is itself a group. Here $(H, *_{|H})$ is short for restricting the map $* : G \times G \to G$ to the subset $H \times H$, and obtain a map to G. That we end up in H is part (2) of 2.9.

4. A subgroup H can also be defined as a subset of G such that the restriction of $*$ to $H \times H$ makes $(H, *_{|H})$ to a group.
Proof: Exercise.

5. Let $(G, *)$ be a group. The fact that a subset H is a subgroup of G is often denoted by $H \leqslant G$.

Definition 2.10. 1. A group $(G, *)$ is called a finite group if G has only a finite number of elements, i.e. $\#G = N < \infty$.

2. For a finite group G, the number of elements of the group is called the order of the group: $ord(G) := \#G$.

Example 2.11. (of a finite group)
We already introduced $(\mathrm{Rot}(2), \circ)$, the group of rotations in the plane around $0 = (0, 0)$. Let N be a natural number, $N > 1$.

$$\mathrm{Rot}_N(2) := \left\{ R_\varphi \in \mathrm{Rot}(2) \mid \exists k \in \mathbb{Z}, \varphi = k \cdot \frac{2\pi}{N} \right\}$$

The subset $\mathrm{Rot}_N(2)$ consists of rotations by angles which are multiples of $\frac{2\pi}{N}$. We consider two rotations R_φ, $R_{\varphi'}$ as equal if they make the same for all points in the plane. In particular: $R_0 = R_{2\pi} = R_{k \cdot 2\pi}$ for $k \in \mathbb{Z}$. We also denote this group by C_N

Proposition 2.12. *$Rot_N(2)$ is a finite subgroup of $Rot(2)$ of order N.*

Proof. To check that $\mathrm{Rot}_N(2)$ is indeed a subgroup (see Definition 2.9) is an easy exercise.

Now set $\varphi_N := \dfrac{2\pi}{N}$.
Then note that R_0, R_{φ_N}, $R_{2\varphi_N}$,..., $R_{(N-1)\varphi_N}$ are all pairwise distinct rotations.
By $R_\varphi = R_{\varphi+k\cdot 2\pi}$ it is clear that these are the only distinct elements of $\mathrm{Rot}_N(2)$. Hence the statement of the proposition.
Furthermore we have that:

$$\begin{aligned} R_{k\cdot\varphi_N} = (R_{\varphi_N})^k &= \underbrace{R_{\varphi_N} \circ \cdots \circ R_{\varphi_N}}_{k-times} \\ (R_{k\cdot\varphi_N})^{-1} &= R_{-k\cdot\varphi_N} = R_{(N-k)\varphi_N} \\ (R_{\varphi_N})^N &= R_0 = id \end{aligned} \tag{1}$$

□

Question: Which groups exists with 1, 2, or 3 elements?

2.2 Cyclic groups

We start with an arbitrary group $(G, \cdot)$ with a unit element e. Let $a \in G$. For multiple products of a with itself we use ($n \in \mathbb{Z}$):

$$a^n := \begin{cases} \overbrace{a \cdot a \cdot \cdots \cdot a}^{n-times}, & n > 0 \\ e, & n = 0 \\ \underbrace{(a^{-1}) \cdot (a^{-1}) \cdot \cdots \cdot (a^{-1})}_{|n|-times} & n < 0 \end{cases}$$

As an exercise, one verifies

$$a^n \cdot a^m = a^{n+m}, \quad (a^n)^{-1} = a^{-n}, \quad (a^n)^k = a^{n\cdot k}.$$

(In complete accordance with the usual rules of exponents for numbers).
Let $\langle a \rangle := \{a^n \mid n \in \mathbb{Z}\}$. Then $\langle a \rangle$ is a subgroup of G, as:

$$\begin{gathered} e = a^0 \in \langle a \rangle \\ a^n, a^m \in \langle a \rangle \Rightarrow a^n \cdot a^m = a^{n+m} \in \langle a \rangle \\ a^n \in \langle a \rangle \Rightarrow (a^n)^{-1} = a^{-n} \in \langle a \rangle. \end{gathered}$$

Moreover $\langle a \rangle$ is a commutative group as $a^n \cdot a^m = a^{n+m} = a^{m+n} = a^m \cdot a^n$.

Definition 2.13. 1. A group G is called a cyclic group if there exists an element $a \in G$ such that $\langle a \rangle = G$ (or equivalently: for all $b \in G, \exists n \in \mathbb{Z}$ such that $b = a^n$).

2. If G is a cyclic group, then every $a \in G$ with $\langle a \rangle = G$ is called a generator of G

Example 2.14.
1. $(\mathbb{Z}, +)$ is a cyclic group. The number +1 is a generator, but -1 is also a generator. All other numbers are not generators.

2.$C_N = \mathrm{Rot}_N(2)$ is a cyclic group, $N \in \mathbb{N}$. R_{φ_N}, the rotation by $\varphi_N = \frac{2\pi}{N}$ is a generator (look at the description of its elements $R_0, R_{\varphi_N}, \ldots, R_{(N-1)\varphi_N}$). If N is a prime number, then every $R_{k\cdot\varphi_N}$ for $k = 1, \ldots, N-1$ $(k \neq 0)$ is a generator.

Up to isomorphy (see in the definition below (2.26)) , $\mathbb{Z}$ and C_N, $N \in \mathbb{N}$ are the only cyclic groups. We have:

$$\#\mathbb{Z} = \infty, \quad \#C_N = N$$

.

In an arbitrary group G (not necessarily cyclic), for an element $a \in G$, the subgroup $\langle a \rangle \leqslant G$ is a cyclic group. It is called the cyclic group generated by a.

Definition 2.15. Given an arbitrary group G.

1. An element $c \in G$ is called an element of finite order if and only if $\exists n \in \mathbb{N}$ $(n \neq 0!)$ such that $c^n = e$. It is called of infinite order if there does not exist such a natural number n.

2. The minimal $n \in \mathbb{N}$ such that $c^n = e$ is called order of c, denoted by $\mathrm{ord}(c) = n$. If there does not exist such an n, one sets $\mathrm{ord}(c) = \infty$.

The order of an element $c \in G$ is equal to the number of elements in $\langle c \rangle$ (i.e. $\mathrm{ord}(c) = \#\langle c \rangle$). Recall that we defined the order of a group as the number of elements it contains. Hence, we can also formulate this statement as:
The order of an element equals the order of the cyclic subgroup it generates.
If we look at the finite order case, i.e $\mathrm{ord}(c) = n$, we see:

$$\langle c \rangle = \{e, c, c^2, c^3, \ldots, c^{n-1}\}$$

Note that $c^n = e$ and $c^{-1} = c^{n-1}$.

2.3 Little Fermat

Theorem 2.16. *(Theorem of Lagrange)*
Let G be a finite group and $H \leqslant G$ a subgroup. Then the order of H divides the order of G. In symbols: $\#H \mid \#G$.

The proof will be given in Appendix C.

We can deduce a very important consequence from this:

Theorem 2.17. *Let G be a finite group of order n. Then every element of G has finite order and its order divides n.*

Proof. **a)** Let $c \in G$ be an element. Consider the $n+1$ elements $(\, c, c^2, c^3, \ldots, c^n, c^{n+1}\,)$. But G has only n different elements. So two of these have to coincide, i.e. $\exists k, l, k \neq l$ (take $k > l$) such that $c^k = c^l$. If we multiply this identity by c^{-l} we get $c^{k-l} = e$. Hence c is of finite order.
b) Now let m be the order of c. By the discussion above we know: $\#\langle c\rangle = \mathrm{ord}(c) = m$. But our Theorem 2.16 says that $\#\langle c\rangle$ divides n. Hence the claim. □

From this theorem we draw an important consequence.

Theorem 2.18. *(Little Fermat)*
Let G be a finite group, $n = \#G$ the order of G. Let $c \in G$ be an arbitrary element. Then $c^n = e$ (or equivalently $c^{n+1} = c$).

Proof. Let m be the order of c. From Theorem 2.17, we know that m divides n, i.e. $\exists k \in \mathbb{N} : n = m \cdot k$. Now $c^n = c^{m\cdot k} = (c^m)^k = e^k = e$, which is the claim. □

As an application, we will show:

Theorem 2.19. *Let G be a finite group of order p ($\#G = p$) , where p is a prime number. Then G is a cyclic (hence commutative) group and every element $c \in G$, $c \neq e$ (the unit element), is a generator of G.*

Proof. Take $a \in G$, then by Theorem 2.17 $\mathrm{ord}(a)$ divides p. But as p is a prime number, only 1 and p will divide it. $\mathrm{ord}(a) = 1$ if and only if $a = e$. Hence for $a \neq e$, $\mathrm{ord}(a) = p$. The cyclic subgroup $\langle a\rangle$ generated by a has p elements. This means $\langle a\rangle = G$. And G is cyclic. □

Remark 4. Why is it so interesting to have generators of a group?
The elements of a group can be very complicated objects. But we want to do calculation with them on the computer. Recall that if we identified a group as a cyclic group and found a generator b, then every element c can be represented by $c = b^n$, with n a number. If we take another element d with representation $d = b^m$, the product $c \cdot d = b^{n+m}$ can now be easily calculated with respect to the generator b by adding the two numbers $n + m$. Also calculating the inverse of c is obtained by taking $-n$ as the exponent. Of course we have to be careful as there is a choice involved, namely the choice of a generator.

In the opposite direction, knowing a generator a of the group and an element b of the group it is usually very hard (if the group and its generator is chosen accordingly) to find the number $x \in \mathbb{N}$ such that $a^x = k$. Determining x is called discrete logarithm problem. The *Diffie-Hellman key exchange scheme* (which is a public-key protocol) is based on the complexity of the discrete logarithm problem.

Example 2.20. The Diffie-Hellman key exchange scheme.

The parties A and B choosse a cyclic group G and a generator a of G. These data can be know to everybody.

The party A chooses a secret key k and calculates a^k. The party B chooses a secret key l and calculate a^l. They exchange these values. Hence A receives a^l and takes its kth power and calculates $(a^l)^k = a^{l \cdot k}$. The party B receives a^k and takes its lth power and calculates $(a^k)^l = a^{l \cdot k}$. Now $a^{l \cdot k}$ is their shared secret. Known (as it went over the channel) are a, a^l, a^k but not $a^{l \cdot k}$. To calculate this one would need to know what l or k is. Otherwise formulated one would need to calculate the discrete logarithm of a^l and/or a^k. Encoding (means calculating a^k) can be done very effectively by squaring and multiplying

2.4 Residue class group

Here we will introduce another realization of the cyclic group of order n.
We consider the set of integers. Let $n \in \mathbb{N}$, $n > 1$, be fixed. We call two numbers $a, b \in \mathbb{Z}$ equivalent modulo n (in symbols $a \underset{n}{\sim} b$) if $b - a$ is divisible by n. In other words, $\exists k \in \mathbb{Z}$ with $b - a = k \cdot n$. Sometimes one also uses "b is congruent to $a \mod n$", respectively $b \equiv a \mod n$.
This notion defines an equivalence relation on the set of integers. This means we have:
1) $\forall a \in \mathbb{Z} \quad a \underset{n}{\sim} a$ (reflexivity).
2) If $a \underset{n}{\sim} b$, then $b \underset{n}{\sim} a$ (symmetry).
3) If $a \underset{n}{\sim} b$ and $b \underset{n}{\sim} c$, then $a \underset{n}{\sim} c$ (transitivity).

Exercise 2.21. Verify these 3 conditions.

For every $a \in \mathbb{Z}$ we introduce its equivalence class

$$\overline{a} := \{b \in \mathbb{Z} \mid b \underset{n}{\sim} a\} =: a \mod n.$$

Both notations $\overline{a}$ and $a \mod n$ are in use.

We have the following result

$$\overline{a} \cap \overline{b} = \begin{cases} \overline{a} & \text{if } b \underset{n}{\sim} a \\ \varnothing & \text{if } b \underset{n}{\not\sim} a \end{cases} . \tag{2}$$

We denote by $\mathbb{Z}_n$ the set of all equivalence classes, i.e.

$$\mathbb{Z}_n = \{\overline{a} \mid a \in \mathbb{Z}\}.$$

Note that the elements in $\mathbb{Z}_n$ are not numbers, but subsets of numbers. Also recall that if we give a set by listing its elements that even if we list an element twice it will be only counted once. For example in $\mathbb{Z}_n$ our list has infinitely many elements as we run through $\mathbb{Z}$. But by the result (2) two elements $\overline{a}$ and $\overline{b}$ are the same if and only if $b \underset{n}{\sim} a$ (or equivalently $(b-a)$ is divisible by n). Hence, a lot of the elements will be the same. In fact this set is finite:

Proposition 2.22.
$\#\mathbb{Z}_n = n$ and $\mathbb{Z}_n = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$ where these elements are distinct.

Proof. Given $a \in \mathbb{Z}$, we can write it after division by n with rest as $a = k \cdot n + r$ with $k \in \mathbb{Z}$ and $0 \leq r < n-1$. Hence $a - r = k \cdot n$, which says $r \underset{n}{\sim} a$, and morevoer $\overline{r} = \overline{a}$. From this follows that it is enough to consider the residue classes $\overline{r}$ of the noted type. In particular $\#\mathbb{Z}_n \leq n$. To show equality, we need to show that none of them coincides with another one. Assume $0 \leq r_1, r_2 < n$ with $\overline{r_1} = \overline{r_2}$, but $r_1 \neq r_2$. We might assume $r_2 > r_1$. From $\overline{r_1} = \overline{r_2}$ it follows $r_2 = k \cdot n + r_1$, hence $r_2 - r_1 = k \cdot n$. From $r_1 \neq r_2$ it follows $k > 0$, which gives $r_2 \geq n$. This is a contradiction. □

Next we want to introduce on $\mathbb{Z}_n$ a group structure coming from the addition in $\mathbb{Z}$. We define:

$$\overline{a} \oplus \overline{b} := \overline{a+b}.$$

The first step is to show that this is well-defined. This means that if $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$, then $\overline{a+b} = \overline{a'+b'}$. $\overline{a} = \overline{a'}$ means $a' = a + k \cdot n$, $\overline{b} = \overline{b'}$ means $b' = b + l \cdot n$. Hence

$a' + b' = a + b + (k + l)n$ and moreover $a' + b' \underset{n}{\sim} a + b$, or $\overline{a + b} = \overline{a' + b'}$.
Next we show that the operation fullfills the axioms of a group structure.

(1) associativity
$(\overline{a} \oplus \overline{b}) \oplus \overline{c} = \overline{(a + b)} \oplus \overline{c} = \overline{(a + b) + c}$
$\overline{a} \oplus (\overline{b} \oplus \overline{c}) = \overline{a} \oplus \overline{(b + c)} = \overline{a + (b + c)}$
The last two expressions are the same as addition in $\mathbb{Z}$ is associative.

(2) The unit is give by $\overline{0}$.

(3) Given $\overline{a}$, then $-\overline{a} = \overline{-a}$.
The conditions (2) and (3) are obvious from the very definition of $\oplus$.

Moreover we have $\overline{2} = \overline{1} \oplus \overline{1} = 2 \cdot \overline{1}, \quad \overline{3} = \overline{1} \oplus \overline{1} \oplus \overline{1} = 3 \cdot \overline{1}, \quad \ldots, \quad \overline{r} = r \cdot \overline{1}.$

Proposition 2.23.
$(\mathbb{Z}_n, \oplus)$ is a (commutative) cyclic group of order n. A generator is given by $\overline{1}$.

Note that as we write our "product" in the group additively, the multiple product c^k corresponds to multiple sums $k \cdot c$.

Warning: $\overline{1}$ is not the only generator. $\overline{n - 1}$ is also a generator. Depending on n there might be even more generators.
In general $\overline{0}$ consists of these integers which are divisible by n.

Example 2.24. The most simple example is given for $n = 2$. Then

$$\mathbb{Z}_2 = \{\overline{0}, \overline{1}\}, \quad \#\mathbb{Z}_2 = 2. \quad \overline{0} = \{k \in \mathbb{Z} \mid k \text{ even}\}, \quad \overline{1} = \{k \in \mathbb{Z} \mid k \text{ odd}\}.$$

Our sum is given by

$$\overline{0} \oplus \overline{0} = \overline{0}, \quad \overline{0} \oplus \overline{1} = \overline{1}, \quad \overline{1} \oplus \overline{0} = \overline{1}, \quad \overline{1} \oplus \overline{1} = \overline{0}.$$

This corresponds to the fact that the sum of two even numbers is even, the sum of an even and an odd number is odd and the sum of two odd numbers is even.

Example 2.25.
1) For $n = 3$ we get $\mathbb{Z}_3 = \{\overline{0}, \overline{1}, \overline{2}\}$. In this case $\overline{1}$ and $\overline{2}$ are generators.

2) For $n = 4$ we get $\mathbb{Z}_4 = \{\overline{0}, \overline{1}, \overline{2}, \overline{3}\}$. $\overline{1}$ and $\overline{3}$ are generators, $\overline{2}$ (and of course $\overline{0}$)

are not.

3) By Theorem 2.19, for $\mathbb{Z}_p$, $p \in \mathbb{P}$ (prime numbers), every element $\overline{r} \in \mathbb{Z}_p$, with $\overline{r} \neq \overline{0}$ will be a generator.

Now we found again for every n a cyclic group $\mathbb{Z}_n$ which has exactly n elements, as it was the case with the groups $\mathrm{Rot}_n(2)$. Both groups consist of completely different elements. But as mathematical groups consisting only of elements and their product structure they should be "the same", as we already mentioned that there is up to "isomorphy" only one cyclic group of order n. Next we want to describe what an isomorphy is in more detail. "Isomorphy" should reflect that both objects have the same "structure". The mathematical term for this is the existence of a homomorphism.

Definition 2.26. Let $(G_1, \cdot_1)$ and $(G_2, \cdot_2)$ be two groups.
A map $\phi : G_1 \to G_2$ is called homomorphism (of groups) if

$$\forall b, c \in G_1 : \quad \phi(b \cdot_1 c) = \phi(b) \cdot_2 \phi(c).$$

If ϕ is additionally bijective it is called an isomorphism.

In words: It doesn't matter if one takes the product before one maps the elements or after one maps them individually.

The existence of such an isomorphism between G_1 and G_2 is indicated by $G_1 \cong G_2$.

One easily verifies from the condition:
a) If e_1 is the unit element is G_1 and e_2 is the unit element in G_2, then $\phi(e_1) = e_2$.
b) $\phi(g^{-1}) = (\phi(g))^{-1}$.

Example 2.27.
1) The map $\phi : \mathbb{Z} \to \mathbb{Z}_n$ (n fixed) given $\quad b \mapsto \phi(b) = \overline{b} \quad$ is a homomorphism.
We have to check whether $\phi(b + c) = \phi(b) \oplus \phi(c)$. This means whether $\overline{b + c} = \overline{b} \oplus \overline{c}$. But this was exactly the definition of $\oplus$. The map is not bijective as e.g $\phi(0) = \phi(n)$, but $0 \neq n$.
As it does not matter whether we take the sum before we take the residue class or after, we will now mainly drop the clumsy notation $\overline{a} \oplus \overline{b}$ and simply write $\overline{a} + \overline{b}$.

2) Let $\mathrm{Rot}_n(2)$ be the cyclic group of rotations by multiples of the angle $\varphi_n = \frac{2\pi}{n}$. We consider the map

$$\phi : \mathrm{Rot}_n(2) \to \mathbb{Z}_n; \qquad R_{k\varphi_n} \mapsto \phi(R_{k\varphi_n}) := \overline{k}.$$

The map ϕ is bijective as this is a 1-1 correspondance. Moreover: $R_{k\varphi_n} \circ R_{l\varphi_n} = R_{(k+l)\varphi_n}$ where the angle has to be calculated modulo 2π. Hence it is a group homomorphism and we might identify both groups. We also write: $\mathrm{Rot}_n(2) \cong \mathbb{Z}_n \cong C_n$ (as groups).

The isomorphism in Example (2) above is a special case of a general fact. Let G and H be cyclic groups of the same order n. Let b be a generator of G and c be a generator of H. Then there exists a unique group isomorphism $\phi : G \to H$ fixed by $\phi(b) = c$. In fact, by requiring $\phi(b) = c$ and the homomorphy property we get $\phi(b^k) = \phi(b)^k = c^k$, $k = 0, 1, \ldots, n-1$. Hence ϕ is uniquely fixed and it is both a bijection and a group homomorphism. Hence an isomorphy.

Remark: In case that we have an isomorphism $\phi : G \to G$ for the same group G we call it an automorphism.

To close this section, we introduce for a group homomorphism $\phi : G_1 \to G_2$

$$\ker \phi := \{g \in G_1 \mid \phi(g) = e_2\} \leqslant G_1,$$

$$\mathrm{Im}\phi := \{h \in G_2 \mid \exists g \in G_1 : h = \phi(g)\} \leqslant G_2.$$

Both are indeed subgroups of G_1 and G_2 respectively. We show this for $\ker \phi$:
1) $\phi(e_1) = e_2 \Rightarrow e_1 \in \ker \phi$.
2) $a, b \in \ker \phi$: $\phi(a \cdot b) = \phi(a) \cdot \phi(b) = e_2 \cdot e_2 = e_2 \Rightarrow a \cdot b \in \ker \phi$.
3) $a \in \ker \phi$: $\phi(a^{-1}) = (\phi(a))^{-1} = e_2^{-1} = e_2 \Rightarrow a^{-1} \in \ker \phi$.

In fact $\ker \phi$ is an even better substructure than a subgroup. It is a normal subgroup (We will not need it in this course - just for completness let me define it).

Definition 2.28. A subgroup $H \leqslant G$ is called normal if and only if $\forall g \in G, \forall h \in H$: $g^{-1} \cdot h \cdot g \in H$.

In the case that G is a commutative group: $g^{-1} \cdot a \cdot g = a \cdot g^{-1} \cdot g = a \cdot e = a$. Hence there is no additional condition. All subgroups of commutative groups are normal.

Exercise 2.29. Verify that $\ker \phi$ is always a normal subgroup of G_1.

2.5 Multiplication tables

As far as it concerns finite groups, classifying all groups of order n up to isomorphy is just a problem of checking finitely many possibilities for the product in the group. Unfortunately, with increasing n the amount of calculations will increase considerably. Nevertheless it is still possible to do so for small n, even by hand. In particular, if one uses some facts which are true in groups. First let $(G, \cdot)$ be a group, written multiplicatively, e the neutral element.

Proposition 2.30. *Let $a, b \in G$. Then there exists exactly one $x \in G$ and one $y \in G$ such that*

$$a \cdot x = b \qquad y \cdot a = b$$

Warning: We do not assume that our group is commutative. This means x and y might be different. The unique solutions are

$$x = a^{-1} \cdot b \quad \text{and} \quad y = b \cdot a^{-1}$$

We can write down the group structure in form of a table. Let G be a finite group, $\#G = n$, with elements $e, a_2, a_3, \ldots, a_n$ in a fixed order ($a_1 = e$). We write down the table

$\cdot$	e	a_2	a_3	$\ldots$	a_k	$\ldots$	a_n
e					|		
a_2					|		
a_3					|		
$\vdots$					|		
a_j	—	—	—	—	$(a_j \cdot a_k)$		
$\vdots$							
a_n							

At position line j and row k we write the element which is given by the product $(a_j \cdot a_k)$. It has to be an element a_m, with $1 \leq m \leq n$.
The group structure is uniquely fixed if we completely fill the table. Of course this cannot be done arbitrary. The resulting scheme should give a group.

1. Requirement: As $a_l \cdot e = a_l = e \cdot a_l$, the first line and row just duplicates the elements. Hence

$\cdot$	e	a_2	a_3	$\dots$	a_n
e	e	a_2	a_3	$\dots$	a_n
a_2	a_2	$\dots$	$\dots$	$\dots$	$\dots$
a_3	a_3	$\dots$	$\dots$	$\dots$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_n	a_n	$\dots$	$\dots$	$\dots$	$\dots$

2. Requirement: In each line and in each row every element has to appear exactly once. This is a consequence of Proposition 2.30. By this the number of possible entries is already reduced. But still not all scheme of entries will define groups as the associativity conditon has to be verified (i.e. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$). This is admittingly a very stupid task to do for a human being.

Example 2.31.
1) $\#G = 1$. Then $G = \{e\}$ and

$\cdot$	e
e	e

This is the trivial group.

2) $\#G = 2$. $G = \{e, a\}$, $a \neq e$. Then by the first rule

$\cdot$	e	a
e	e	a
a	a	?

By the 2nd rule for "?" only e is possible. Hence

$\cdot$	e	a
e	e	a
a	a	e

In particular, $a^{-1} = a$, ord(a)=2 and $G \cong C_2$, the cyclic group of 2 elements.

3) $\#G = 3$. Of course from the theory we know that G is isomorphic to the cyclic group with 3 elements. Hence there is only one isomorphy which we know already. But as an illustration we want to show this directly. Let us denote the elements by (e, a, b).

$\cdot$	e	a	b
e	e	a	b
a	a	(1)	(2)
b	b	.	.

Next we want to see what elements can be put for (1). By checking row 2, we see that $(1) = a$ is not possible anymore. If we try $(1) = e$, then (2) has to be b, which is in contradiction with row 3 (we would get twice b). Hence necessarily $(1) = b$.

$\cdot$	e	a	b
e	e	a	b
a	a	b	(2)
b	b	(3)	(4)

Now $(2) = (3) = e$ is necessary, which leaves us the only solution $(4) = a$.
Hence a unique table.

4) $\#G = 4$. Here the situation starts to become more interesting.
First, we have the cyclic group C_4. Let us denote the elements by $(e, a, b = a^2, c = a^3)$.

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

This table follows from $a^l \cdot a^m = a^{l+m}$.
The question remains: does there also exist another isomorphy type?
If G has an element of order 4 then necessarily $G \cong C_4$. Hence if there is another isomorphy type the order of the elements are either 1 or 2 (as the orders of the elements have to divide 4).
$\text{ord}(a) = 1 \quad \Leftrightarrow \quad a = e$.
If $\text{ord}(a)=2$, then $a \cdot a = e$, or equivalently $a^{-1} = a$. We write down as much as we know now. (e, a, b, c) are the elements.

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	(1)	(2)
b	b	(3)	e	(4)
c	c	(5)	(6)	e

The elements (1) to (6) are not yet known for the moment. But if we check the second line, then for (1) only c is possible ($(1) = b$ would give a contradiction for row 3). Hence $(2) = b \Rightarrow (4) = a \Rightarrow (3) = c \Rightarrow (5) = b \Rightarrow (6) = a$. In total

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

To show that this defines indeed a group structure one would need to check all associativity relations. But again one could avoid this by remarking that there must be a group G which is given as the product $C_2 \times C_2$ and has exactly 4 elements, hence it is this type. Here we found exactly two different isomorphy types of groups of order 4 (See Exercise 2.32 and 2.33).

Remarks
a) If we change the role of the elements a, b, c, this corresponds to an isomorphism. Hence it does not define another isomorphy type.
b) A group G is a commutative group if and only if the product table is symmetric with respect to the counter diagonal.

$\cdot$	e	a	b	c
e	$\ddots$			
a		$\ddots$		
b			$\ddots$	
c				$\ddots$

Obviously this is the case for our groups of order 4.

5) $\#G = 5$ is again an order which is a prime number. Hence, G is isomorphic to the (commutative) cyclic group of order 5. Hence, only one isomorphy type.

6) $\#G = 6$. Here we will obtain for the first time also a group which is not commutative. Recall that C_6 is a commutative cyclic group and that S_3 has also 6 elements. S_3 is the group of permutations of the letters $\{1, 2, 3\}$.

A challenging exercise is that these two groups are the only isomorphy types of groups of order 6. For this exercise the following facts might be useful:
a) If U and V are two subgroups of G, then $U \cap V$ is also a subgroup of G (exercise).
b) Let U, V be two finite subgroups of a group G and suppose $\gcd(\#U, \#V)=1$. Then $U \cap V = \{e\}$ (exercise).
c) Recall that $\forall a, b \in G$: $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$.
Now make a case distinction:
First case: the group is commutative. Show that then $G \cong C_6$.
Second case: the group is non-commutative. Show that there is only one isomorphy type which in this case has to be S_3.

2.6 Exercises

Exercise 2.32.
Let $(G_1, \cdot_1)$ and $(G_2, \cdot_2)$ be two groups. Define for the Cartesian product $G_1 \times G_2$ the multiplication: $a_1, b_1 \in G_1$, $a_2, b_2 \in G_2$

$$(a_1, a_2) \cdot (b_1, b_2) := (a_1 \cdot b_1, a_2 \cdot b_2).$$

Show that this defines a group structure.

Exercise 2.33.
Let G_1 and G_2 be cyclic groups, $\#G_1 = n_1$, $\#G_2 = n_2$.
a) Assume that $\gcd(n_1, n_2) = 1$. Show that $G_1 \times G_2$ is also a cyclic group and that its order is $n_1 \cdot n_2$.
b) Is the same also true if $\gcd(n_1, n_2) \neq 1$? (Look at $C_2 \times C_2$. Is this a cyclic group? What are the orders of its elements?)

Exercise 2.34.
Let G be a group and U, V subgroups of G.
a) Show that $U \cap V$ is also a subgroup of G.
b) Assume that U and V are finite groups such that $\gcd(\#U, \#V) = 1$ (gcd = greatest common divisor). Show that then $U \cap V = \{e\}$, the trivial group.

Exercise 2.35.
Let $(G, \cdot)$ be a group. Show the following
a) For $a, b, c \in G$ the relation $a \cdot b = a \cdot c$ implies $b = c$.
b) For $a, b \in G$ we have $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$.

Exercise 2.36.
a) Let $Mat(2 \times 2, \mathbb{R})$ be the set of 2×2 matrices with real entries. Take as operation the multiplication of matrices. Show by exhibiting a counter example that in this case $A \cdot B = A \cdot C$ does not necessarily imply $B = C$.
b) Consider the subset of matrices A with $\det A \neq 0$. Show that this subset is a group. Hence for such matrices $A \cdot B = A \cdot C$ implies $B = C$.

Exercise 2.37.
Go again through the classification of the isomorphy type of all groups of order less or equal 5. Are these groups abelian?

Exercise 2.38.
Let S_3 be the group of bijective maps from $\{1, 2, 3\}$ to $\{1, 2, 3\}$. Give all elements of this group. Calculate the order of the elements and find all subgroups of S_3. Is this group abelian?

Exercise 2.39.
Let G be a cyclic group and H a subgroup of G. Show that H is also a cyclic group.

3 Rings and fields

3.1 Basic definitions

For the usual numbers, like integers $\mathbb{Z}$, rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, we have two operations: the sum $+$ and the multiplication $\cdot$; and certain properties relating them like $(a+b)\cdot c = a\cdot c + b\cdot c$ (distributive laws). Such structures are called rings.

Definition 3.1. A set R with two operations

$$+ : R \times R \to R; \qquad (a,b) \mapsto a+b$$

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

is called a ring if and only if

1. $(R,+)$ is a commutative group (we denote the neutral element by 0).
2. $(R,\cdot)$ is an associative composition which has a unit (which we denote by 1), i.e. $1\cdot a = a, \forall a \in \mathbb{R}$.
3. We have the distributive laws

$$\forall a,b,c \in R: \qquad (a+b)\cdot c = a\cdot c + b\cdot c, \quad a\cdot(b+c) = a\cdot b + a\cdot c.$$

The ring is called a commutative ring if $(R,\cdot)$ is commutative.

If needed, we denote the ring by $(R,+,\cdot)$. Examples are $(\mathbb{Z},+,\cdot)$, $(\mathbb{Q},+,\cdot)$, $(\mathbb{R},+,\cdot)$. If we know that $(R,+,\cdot)$ is a ring, then certain properties follow directly from the definition.

Proposition 3.2. *Let $(R,+,\cdot)$ be a ring. Then*

1. $0\cdot a = 0, \quad \forall a \in R.$
2. $(-a)\cdot b = -(a\cdot b), \quad \forall a,b \in R.$
3. $(-a)\cdot(-b) = a\cdot b, \quad \forall a,b \in R.$

Proof. ad1) $0\cdot a + a = 0\cdot a + 1\cdot a = (0+1)\cdot a = 1\cdot a = a \quad \Rightarrow 0\cdot a = 0.$

ad2) $a\cdot b + (-a)\cdot b = (a+(-a))\cdot b = 0\cdot b = 0 \quad \Rightarrow -(a\cdot b) = (-a)\cdot b.$

ad3) $(-a)\cdot b+(-a)\cdot(-b)=(-a)\cdot(b+(-b))=(-a)\cdot 0=0$
$\Rightarrow (-a)\cdot(-b)=-\left((-a)\cdot b\right)=-\left(-(a\cdot b)\right)=a\cdot b$

□

We call an element r in a ring R invertible if $\exists s\in R$: $r\cdot s=1$ and $s\cdot r=1$. We set

$$R^*:=\{r\in R\mid r \text{ is invertible}\}\subseteq R.$$

Proposition 3.3. *R^* is a group, called the group of units of R.*

Proof. a) $1\in R^*$, as $1\cdot 1=1$ and hence 1 is invertible.

b) $r\in R^*\Rightarrow r^{-1}\in R^*$, as: $r\cdot r^{-1}=r^{-1}\cdot r=1$ and we have $(r^{-1})^{-1}=r$.

c) Let $r_1,r_2\in R^*$. We have to show $r_1\cdot r_2\in R^*$.
$r_i\in R^* \quad\Rightarrow\quad \exists r_i^{-1}\in R^* \quad i=1,2.$
We calculate $(r_1\cdot r_2)\cdot(r_2^{-1}r_1^{-1})=r_1\cdot 1\cdot r_1^{-1}=r_1\cdot r_1^{-1}=1$. Also $(r_2^{-1}r_1^{-1})\cdot(r_1\cdot r_2)=1$.
Hence $r_1\cdot r_2$ is invertible $\quad\rightarrow\quad r_1\cdot r_2\in R^*$.

d) Associativity of R^* follows from the associativity of $(R,\cdot)$.

□

Example 3.4. $\mathbb{Z}^*=\{+1,-1\}$ (Exercise); $\quad\mathbb{Q}^*=\mathbb{Q}\setminus\{0\}$.

Definition 3.5. $(\mathbb{K},+,\cdot)$ is a field if

1. $(\mathbb{K},+,\cdot)$ is a commutative ring with $1\neq 0$.
2. $(\mathbb{K}\setminus\{0\},\cdot)$ is a group.

Equivalently to 2., we could also use

$$2'.\quad \mathbb{K}^*=\mathbb{K}\setminus\{0\}.$$

Examples of a field are $(\mathbb{Q},+,\cdot)$, $(\mathbb{R},+,\cdot)$, $(\mathbb{C},+,\cdot)$, whereas $(\mathbb{Z},+,\cdot)$ is not a field, only a ring.

As for groups, we can now consider maps between rings which respect the ring structures.

Definition 3.6. Given $(R_1, +_1, \cdot_1)$ and $(R_2, +_2, \cdot_2)$ two rings. A map $\varphi : R_1 \to R_2$ is called a ring homomorphism if

1. $\varphi(r_1 +_1 s_1) = \varphi(r_1) +_2 \varphi(s_1), \quad \forall r_1, s_1 \in R_1.$
2. $\varphi(r_1 \cdot_1 s_1) = \varphi(r_1) \cdot_2 \varphi(s_1), \quad \forall r_1, s_1 \in R_1.$
3. $\varphi(1_1) = 1_2.$

Examples of rings are given by the quadratic $n \times n$ matrices with entries in a field (or even ring) with matrix addition and matrix multiplication as operations. (*Show this as an exercise!*).

3.2 Residue class rings

Let $n \in \mathbb{N}$ be fixed, $n > 1$. We consider now the ring $(\mathbb{Z}_n, +, \cdot)$, where $\mathbb{Z}_n$ is the set of residue classes mod n (see the previous chapter), with $+$ the addition introduced there, i.e.

$$\overline{a} + \overline{b} := \overline{a + b}.$$

In the similar spirit we introduce the multiplication

$$\overline{a} \cdot \overline{b} := \overline{a \cdot b}.$$

1) The multiplication is well defined: let $\overline{a'} = \overline{a}$, i.e. $a' = a + l \cdot n, \quad k \in \mathbb{Z}$ and $\overline{b'} = \overline{b}$, i.e. $b' = b + l \cdot n, \quad l \in \mathbb{Z}$.

$$\begin{aligned} \overline{a'} \cdot \overline{b'} &= \overline{a' \cdot b'} = \overline{(a + k \cdot n)(b + l \cdot n)} \\ &= \overline{a \cdot b + k \cdot b \cdot n + a \cdot l \cdot n + k \cdot l \cdot n^2} \\ &= \overline{a \cdot b + r \cdot n}, \qquad \text{with a suitable } r \\ &= \overline{a \cdot b} \end{aligned}$$

2) $\overline{1}$ is the unit of multiplication

$$\overline{1} \cdot \overline{a} = \overline{1 \cdot a} = \overline{a} = \overline{a} \cdot \overline{1}.$$

3) The multiplication is commutative

$$\overline{a} \cdot \overline{b} = \overline{a \cdot b} = \overline{b \cdot a} = \overline{b} \cdot \overline{a}.$$

4) The multiplication is distributive

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

We have shown

Proposition 3.7. $(\mathbb{Z}_n, +, \cdot)$ *is a commutative ring.*

The question is: Is $(\mathbb{Z}_n, +, \cdot)$ maybe even a field? To answer this, we first have

Proposition 3.8. *In a field* $\mathbb{K}$*, for* $a, b \in \mathbb{K}$*, with* $a \cdot b = 0$ *it follows that* $a = 0$ *or* $b = 0$*.*

Proof. Assume $a \cdot b = 0$ and $a \neq 0$. Then $\exists a^{-1}$ with $a^{-1} \cdot a = 1$. We multiply the relation $a \cdot b = 0$ by a^{-1} from the left: $a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0$. On the left hand side we have: $a^{-1} \cdot (a \cdot b) = (a^{-1} \cdot a) \cdot b = 1 \cdot b = b$. Hence $b = 0$. □

Now we consider $(\mathbb{Z}_n, +, \cdot)$ for $n = 6$. Then $\overline{3}, \overline{2} \neq 0$, but $\overline{3} \cdot \overline{2} = \overline{6} = 0$. Hence $(\mathbb{Z}_6, +, \cdot)$ cannot be a field.

Theorem 3.9. *Let* $n \in \mathbb{N}$*,* $n > 1$*.*
Then $(\mathbb{Z}_n, +, \cdot)$ *is a field if and only if* n *is a prime number* ($n \in \mathbb{P}$*).*

Before we can prove the theorem, we have to recall some properties for the integers $\mathbb{Z}$, respectively for the positive integers $\mathbb{N}$.
Let $a, b \in \mathbb{N}$, then b is called a divisor of a if and only if there exists $c \in \mathbb{N}$ such that $b \cdot c = a$. We write $b \mid a$. Given $a, b \in \mathbb{N}$, then $c \in \mathbb{N}$ is called a divisor of a and b if $c \mid a$ and $c \mid b$.
A number $c \in \mathbb{N}$ is called greatest common divisor of a and b (denoted by $c = \gcd(a, b)$) if $c \mid a$ and $c \mid b$, and if $d \mid a$ and $d \mid b$ then $d \mid c$. For sure 1 is always a common divisor of a and b, but not necessarily the $\gcd(a, b)$. Some examples: $1 = \gcd(11, 13)$, $1 = \gcd(2, 3)$, $2 = \gcd(6, 8)$, $4 = \gcd(4, 16)$. The greatest common divisor always exists. We will return to it further down in Section 3.3.
For the moment we only quote the following

Lemma 3.10. *(from number theory)*
Given two numbers $k, n \in \mathbb{N}$. Then the gcd(k, n) is an integer combination of k and n:

$$\exists s_1, s_2 \in \mathbb{Z}: \quad gcd(k, n) = s_1 \cdot k + s_2 \cdot n. \tag{3}$$

Moreover, the integers s_1 and s_2 can be effectively calculated.

Proof. (of Theorem 3.9).

$\Rightarrow$: Assume that $\mathbb{Z}_n$ is a field and $n \in \mathbb{N}$ is not a prime. Then $\exists k, l \in \mathbb{N}: \quad n = k \cdot l, \quad 1 < k, l < n$. But this yields $\overline{0} = \overline{n} = \overline{k} \cdot \overline{l}$, with $\overline{k}, \overline{l} \neq \overline{0}$. Hence $\mathbb{Z}_n$ cannot be a field.

$\Leftarrow$: Let $n \in \mathbb{P}$. We have to show that for $\overline{k} \neq \overline{0}$, there exists an $\overline{l} \neq \overline{0}$ such that $\overline{k} \cdot \overline{l} = \overline{1}$. $\overline{k} \neq \overline{0}$ says that $n \nmid k$. As n is a prime, $\gcd(n, k) = 1$. Hence by the statement of Lemma 3.10: $1 = s \cdot n + t \cdot k$ with $s, t \in \mathbb{Z}$. Now we consider the residue class: $\overline{1} = \overline{s} \cdot \overline{n} + \overline{t} \cdot \overline{k} = \overline{s} \cdot \overline{0} + \overline{t} \cdot \overline{k} = \overline{t} \cdot \overline{k}$. In other words, $l = (\overline{k})^{-1}$ and every $\overline{k} \neq \overline{0}$ is invertible.

□

Example 3.11. (for calculating the inverse elements)
n=3:
$\overline{k} = \overline{1} \quad (\overline{k})^{-1} = \overline{1};$
$\overline{k} = \overline{2} \quad (\overline{k})^{-1} = \overline{2}, \quad$ as $\overline{2} \cdot \overline{2} = \overline{4} = \overline{1};$

n=5:
$\overline{k} = \overline{1} \quad (\overline{k})^{-1} = \overline{1};$
$\overline{k} = \overline{2} \quad (\overline{k})^{-1} = \overline{3}, \quad$ as $\overline{2} \cdot \overline{3} = \overline{6} = \overline{1};$
$\overline{k} = \overline{3} \quad (\overline{k})^{-1} = \overline{2};$
$\overline{k} = \overline{4} \quad (\overline{k})^{-1} = \overline{4}, \quad$ as $\overline{4} \cdot \overline{4} = \overline{16} = \overline{1};$

Here we just guessed the inverse element. But Equation 3 will give us a systematic method for calculating the inverse (the extended Euclid algorithm).

The method of the proof of Theorem 3.9 can be extended to prove

Theorem 3.12. *Given $(\mathbb{Z}_n, +, \cdot)$, $n \in \mathbb{N}$, $n > 1$. Then*

$$\mathbb{Z}_n^* = \{\overline{k} \mid gcd(k, n) = 1\}.$$

Proof. With the same argument as above: if $\gcd(k,n)=1$, then $\overline{k}$ is invertible. Conversely, if $\overline{k}$ is invertible, this implies $\overline{k}\cdot\overline{s}=\overline{1}$, with $s\in\mathbb{N}$, or $1=k\cdot s+m\cdot n$, with $m\in\mathbb{Z}$. If l is a divisor of k and n, then l is also a divisor of $k\cdot s+m\cdot n=1$. But the only divisor of 1 is 1. Hence, 1 is indeed the $\gcd(k,n)$. □

An important function of number theory is the Euler φ - function defined by

$$\varphi(n) := \#\mathbb{Z}_n^* = \{k\in\mathbb{N} \mid \gcd(k,n)=1, \quad 1\le k<n\}.$$

For $p\in\mathbb{P}$ a prime number, the notation $\mathbb{F}_p$ is a short form for the residue class ring $(\mathbb{Z}_p,+,\cdot)$, which by Theorem 3.9 is a field. In contrast to the well-known fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, it contains only finitely many elements (in fact exactly p elements).

Definition 3.13. Given a field $\mathbb{K}$ with multiplicative unit 1. If there is an $n\in\mathbb{N}$ such that

$$n\cdot 1=\underbrace{1+1+\ldots+1}_{n \text{ times}}=0$$

then we say that $\mathbb{K}$ is of finite characteristics. The smallest $n\in\mathbb{N}$ such that $n\cdot 1=0$ is called the characteristics of $\mathbb{K}$, char $\mathbb{K}$. If there does not exist such an n then we put char $\mathbb{K}:=0$.

Remark 5.
a) If $\mathbb{K}$ is a finite field then the characteristic is always finite. But attention: there exists fields with infinitely many elements which have finite characteristics.

b) $\mathrm{char}(\mathbb{F}_p)=p$. As here $n\cdot\overline{1}=\underbrace{\overline{1}+\overline{1}+\ldots+\overline{1}}_{n \text{ times}}=\overline{n}$ and $\overline{n}=\overline{0}$ if and only if n is a multiple of p.

c) $\mathrm{char}(\mathbb{Q})=\mathrm{char}(\mathbb{R})=\mathrm{char}(\mathbb{C})=0$.

Proposition 3.14. *Let $n=$ char $\mathbb{K}$, $n\neq 0$. Then n is a prime number ($n\in\mathbb{P}$).*

Proof. Assume $n \notin \mathbb{P}$, then $n = p \cdot q$ with $p, q \in \mathbb{N}$, $1 < p, q < n$. Let $x = p \cdot 1$, $y = q \cdot 1$. Then

$$x \cdot y = \underbrace{(1 + \ldots + 1)}_{p \text{ times}} \cdot \underbrace{(1 + \ldots + 1)}_{q \text{ times}} = \underbrace{1 \cdot 1 + \ldots + 1 \cdot 1}_{p \cdot q \text{ times}} = \underbrace{1 + \ldots + 1}_{p \cdot q \text{ times}} = 0$$

But as we are in a field, $x \cdot y = 0$ is only possible if $x = 0$ or $y = 0$; hence, e.g. $p \cdot 1 = 0$. This is a contradiction as the characteristic is the smallest number such that $n \cdot 1 = 0$. It follows that n cannot be decomposed. □

For the moment without a proof, we have

Theorem 3.15. *1. Let $\mathbb{K}$ be a finite field which has q elements. Then there exists a prime number $p \in \mathbb{P}$, and a natural number $n \in \mathbb{N}$ such that $q = p^n$ and char $\mathbb{K} = p$.*

2. Moreover, for every $p \in \mathbb{P}$ and $n \in \mathbb{N}$ there exists up to isomorphy one and only one field with $q = p^n$ elements. These fields are denoted by $\mathbb{F}_q$ respectively $\mathbb{F}_{p^n}$ and are sometimes called Galois fields. Attention: for $n > 1$: $\mathbb{F}_{p^n} \neq \mathbb{Z}_{p^n}$ as rings.

These finite fields are of fundamental importance in coding theory and cryptography. Binary codes are directly related to $\mathbb{F}_{2^n}$. In a later chapter, we will construct these fields $\mathbb{F}_{p^n}$ and we will deduce how the calculation is done in these fields. Note that in $\mathbb{F}_{p^n}$ there is always a subfield isomorphic to $\mathbb{F}_p$. It is realized by assigning $\overline{k} \mapsto k \cdot 1$ (1 is the one in $\mathbb{F}_{p^n}$).

If we take $a \in \mathbb{F}_{p^n}$, $a \neq 0$, then $a \in \mathbb{F}_{p^n} \setminus \{0\}$ which is a group under multiplication. It has $p^n - 1$ elements. Little Fermat shows $a^{p^n - 1} = 1$. This multiplied by a gives

$$a^{p^n} = a, \quad \forall a \in \mathbb{F}_{p^n}$$

as this is also true for $a = 0$.

Again without proof, we have

Theorem 3.16. *The multiplicative group $\mathbb{F}_{p^n}^*$ is a cyclic group.*

Recall that if we have given a generator of a cyclic group, all calculations in the group can be easily realized on a computer. This is rather helpful for crypto-schemes. As for a cyclic group all subgroups are also cyclic (Exercise). This can be also employed for the subgroups of the multiplicative group.

3.3 Greatest common divisor and Euclid algorithm

Given $m, n \in \mathbb{N}$, recall that the greatest common divisor $d = \gcd(m, n)$ was defined to be a divisor of m and n, such that every common divisor of m and n also divides d. Clearly, we can decompose m and n into products of primes

$$n = p_1^{i_1} \cdot \ldots \cdot p_l^{i_l}, \qquad m = p_1^{j_1} \cdot \ldots \cdot p_l^{j_l}.$$

(Here we allow the exponents to be equal to zero in order to write down the same primes for m and n.) Then

$$\gcd(n, m) = p_1^{\min(i_1, j_1)} \cdots p_l^{\min(i_l, j_l)}.$$

But to use this formula we first need the prime decomposition of n and m. Unfortunately this is a very hard problem for large numbers n and m. In fact, no polynomial time algorithm is known (in the length of representations for n). Hence it is not a reasonable method. But there is another very effective method: the algorithm of Euclid.

First recall that we have for the integers $\mathbb{Z}$ the division with rest, i.e. given $m, n \in \mathbb{N}$, then there exists unique $q, r \in \mathbb{N}$ such that

$$m = q \cdot n + r, \qquad \text{with } 0 \leq r < n.$$

The number m is divisible by n if and only if the rest $r = 0$.

We want to apply now this division successively to find the $\gcd(m, n)$. After changing the role of m and n, we might assume $m \geq n$.

1. step:
$m = q_1 \cdot n + r_1, \qquad 0 \leq r_1 < n.$
If $r_1 = 0$, then $n \mid m$. Hence $\gcd(m, n) = n$ and we are done. Otherwise we continue

2. step:
$n = q_2 \cdot r_1 + r_2, \qquad 0 \leq r_2 < r_1.$
We stop if $r_2 = 0$ If not then:

3. step:
$r_1 = q_3 \cdot r_2 + r_3, \qquad 0 \leq r_3 < r_2,$
and so on: $r_{k-3} = q_{k-1} \cdot r_{k-2} + r_{k-1},$
$r_{k-2} = q_k \cdot r_{k-1} + r_k,$
$r_{k-1} = q_{k+1} \cdot r_k + 0.$

As in each step the rest r_l is either zero and then we stop or will get smaller but still remains ≥ 0. This process has to stop after finitely many steps. Let us assume

that in the step k, $r_k \neq 0$ but in step $k+1$, $r_{k+1} = 0$.

Claim
$d = \gcd(n, m) = r_k$ and $\exists r, s \in \mathbb{Z}$ such that $d = r \cdot n + s \cdot m$.

For this we invert the chain of formulas above:
$r_k = -q_k \cdot r_{k-1} + r_{k-2}$,
$r_{k-1} = -q_{k-1} \cdot r_{k-2} + r_{k-3}$,
$\vdots$
$r_3 = -q_3 \cdot r_2 + r_1$,
$r_2 = -q_2 \cdot r_1 + n$,
$r_1 = -q_1 \cdot n + m$.
From the last equation it follows that r_1 is an integer combination of n and m. Hence the 2^{nd} last equation shows that r_2 is also an integer combination of n and m. And so on, and finally r_k is an integer combination of n and m.
Next we show that r_k divides n and m. From the last equation in the original algorithm it follows that $r_k \mid r_{k-1}$. Hence $r_k \mid (q_k \cdot r_{k-1} + r_k) = r_{k-2}$ and so on until finally we have $r_k \mid n$ and $r_k \mid m$. If d is any divisor of n and m, then d divides any integer combination of n and m. In particular also r_k. But this implies that r_k is the $\gcd(n, m)$.

This algorithm is extremely effective as the size of the rest goes down very fast. Furthermore it easily allows to calculate the integer coefficients r and s. We just use the backward chain of formulas

$$r_k = -q_k \cdot r_{k-1} + r_{k-2}$$

and plug in the formula for r_{k-1}

$$r_k = -q_k(-q_{k-1} \cdot r_{k-2} + r_{k-3}) + r_{k-2} = (q_k \cdot q_{k-1} + 1)r_{k-2} - q_k \cdot r_{k-3}.$$

Now continue with r_{k-2} and so on until all the rests are replaced finally by n and m. This supplies a proof of Lemma 3.10.

Example 3.17. As an example, we calculate the multiplicative inverse element of $\overline{5}$ in $\mathbb{F}_{23} = (\mathbb{Z}_{23}, +, \cdot)$. Of course $1 = \gcd(5, 23)$ as 23 is a prime number. But we are interested in the r and s such that

$$1 = r \cdot 5 + s \cdot 23.$$

From $\overline{1} = \overline{r} \cdot \overline{5} + \overline{s} \cdot \overline{23}$, with as $\overline{23} = \overline{0}$, it follows $(\overline{5})^{-1} = \overline{r}$. We start

$$\underline{23} = 4 \cdot \underline{5} + 3, \qquad 3 = \underline{23} - 4 \cdot \underline{5}$$

$$\underline{5} = 1 \cdot \underline{3} + 2, \qquad 2 = \underline{5} - 1 \cdot \underline{3}$$
$$\underline{3} = 1 \cdot \underline{2} + 1, \qquad 1 = \underline{3} - 1 \cdot \underline{2}$$

1 is the greatest common divisor. Hence we get

$$1 = \underline{3} - 1 \cdot \underline{2} = \underline{3} - 1 \cdot (\underline{5} - 1 \cdot \underline{3}) = 2 \cdot \underline{3} - 1 \cdot \underline{5} = 2 \cdot (\underline{23} - 4 \cdot \underline{5}) - 1 \cdot \underline{5} = 2 \cdot \underline{23} + (-9) \cdot \underline{5}.$$

Hence: $(\overline{5})^{-1} = \overline{-9} = \overline{23 - 9} = \overline{14}$.
(A short verification: $\overline{5} \cdot \overline{14} = \overline{70} = \overline{3 \cdot 23 + 1} = \overline{1}$.)

3.4 Exercises

Exercise 3.18.
Let $\varphi(n)$ be Euler's phi-function which is defined as the number of elements in $\mathbb{Z}_n^*$, the group of units in the ring $\mathbb{Z}_n$. Show that for p and q prime, one has

$$\varphi_p = p - 1, \qquad \varphi_{(p \cdot q)} = (p - 1)(q - 1).$$

Exercise 3.19.
Let $\mathbb{F}_{23}$ be the finite field with 23 elements. For $\overline{7} = 7 \mod 23$ calculate the inverse element with respect to the multiplication by using Euclid's algorithm (e.g. via expressing the $\gcd(n, m)$ as integer combination of n and m).

Exercise 3.20.
Calculate the inverse elements of the following elements, if they exist.

$$\overline{5} \mod 3, \qquad \overline{5} \mod 9, \qquad \overline{5} \mod 11,$$
$$\overline{27} \mod 11, \qquad \overline{27} \mod 64, \qquad \overline{27} \mod 81, \qquad \overline{27} \mod 101.$$

Exercise 3.21.
a) As it is known from the lecture, $\mathbb{F}_7^*$ is a cyclic group. Determine all generators.
b) An element a of a field $\mathbb{K}$ is called a square if there exists b in $\mathbb{K}$ such that $b^2 = a$. Determine all squares in $\mathbb{F}_7^*$.

4 The ring of polynomials

4.1 Basic definitions

Here we will repeat the construction of the polynomial ring in one variable over a field $\mathbb{K}$. From the mathematical point of view there are different ways to introduce polynomials. Here we do it in the most intuitive way. But a warning is in order: polynomials are not the same as polynomial functions.

We start with a formal symbol X and consider associated symbols X^k, $k \in \mathbb{N}_0$, i.e. $B := \{X^0, X^1, X^2, \ldots, X^k, \ldots\}$. The elements in B are called monomials. One way to introduce polynomials over a field $\mathbb{K}$ is to say the space of polynomials $\mathbb{K}[X]$ is the vector space over $\mathbb{K}$ with basis B. In a more direct way one can say that a polynomial p over $\mathbb{K}$ is a formal expression

$$p(X) := a_0X^0 + a_1X^1 + \ldots + a_nX^n = {\sum_{k\in\mathbb{N}_0}}' a_kX^k \tag{4}$$

with $a_k \in \mathbb{K}$ and $\sum'$ denotes that only finitely many of the a_k are different from zero. Quite often we drop the ' in the notation. If we introduce them as formal expression, we have to identify two formal expressions when they coincide up to a reordering of the terms or they coincide up to terms with coefficients zero.
For example the following polynomials are the same

$$1 \cdot X + 1 \cdot X^0 \text{ and } 1 \cdot X^0 + 1 \cdot X,$$

$$1 \cdot X^0 + 0 \cdot X^1 + 1 \cdot X^2 \text{ and } 1 \cdot X^2 + 0 \cdot X^5 + 1 \cdot X^0.$$

For simplicity we also write

$$1 \cdot X^k = X^k, \quad 1 \cdot X^0 = 1, \quad X^1 = X.$$

The set of polynomials is denoted by $\mathbb{K}[X]$. Given two polynomials $f(X) = \sum_{k\in\mathbb{N}_0} a_kX^k$, $g(X) = \sum_{k\in\mathbb{N}_0} b_kX^k$. Then the sum is defined to be

$$(f+g)(X) := \sum_{k\in\mathbb{N}_0} (a_k + b_k)X^k.$$

Next we introduce the product of two polynomials. We start with the definitions for the monomials:

$$X^k \cdot X^l := X^{k+l}.$$

With this definition we obtain

$$(X)^k = \underbrace{X \cdot X \cdots X}_{k \text{ times}} = X^k.$$

We extend the multiplication of the monomials to the polynomials in a $\mathbb{K}$-bilinear manner. This means

$$(f \cdot g)(X) := \left(\sum_k a_k X^k\right) \cdot \left(\sum_l b_l X^k\right) = \sum_{k,l} a_k \cdot b_l \cdot X^k \cdot X^l = \sum_{k,l} a_k b_l X^{k+l}.$$

Reordering into powers of X gives

$$(f \cdot g)(X) = \sum_m \left(\sum_{n=0}^{m} a_n \cdot b_{m-n}\right) X^m.$$

Proposition 4.1. $(\mathbb{K}[X], +, \cdot)$ *is a commutative ring.*

The details are left as an exercise. The neutral element for the addition is the zero polynomial $0 = 0 \cdot X^0 + 0 \cdot X^1 + \ldots$. For the polynomial $f(x) = \sum a_k X^k$, the inverse element with respect to addition is $(-f)(x) = \sum(-a_k)X^k$. The neutral element with respect to multiplication is $1 = 1 \cdot X^0 + 0 \cdot X^1 + \ldots$. From the mathematical point of view it is not necessary to put always the formal symbol X to f. We sometimes denote a polynomial $f(X)$ simply by f.

Definition 4.2. Given a polynomial $f \in \mathbb{K}[X]$, $f \neq 0$ with $f(X) = \sum_k a_k X^k$. Then its degree is defined as

$$\deg f := \max\{k \in \mathbb{N}_0 \mid a_k \neq 0\}.$$

For consistency we set $\deg(0) = -\infty$.

Example 4.3. $f(X) = X^3 - X$. We have $\deg f = 3$.

Proposition 4.4. *Let* $f, g \in \mathbb{K}[X]$.

1. $\deg(f + g) \leq \max(\deg f, \deg g)$.
2. *If* $\deg f \neq \deg g$, *then* $\deg(f + g) = \max(\deg f, \deg g)$.
3. $\deg(f \cdot g) = \deg f + \deg g$.

The verification is an exercise.

By assigning to $\alpha \in \mathbb{K} \mapsto \alpha \cdot X^0 \in \mathbb{K}[X]$, we define a map $\mathbb{K} \to \mathbb{K}[X]$ which is injective (an embedding) and a ring homomorphism. The image of $\mathbb{K} \setminus \{0\}$ consists exactly of the polynomials of degree zero.

Proposition 4.5. *The group of multiplicatively invertible elements $\mathbb{K}[X]^*$ of the polynomial rings equals $\mathbb{K}^* = \mathbb{K} \setminus \{0\}$.*

Proof. Recall that $f \in \mathbb{K}[X]^*$ if and only if there exists $g \in \mathbb{K}[X]$ such that $f \cdot g = 1$. This implies $\deg(f \cdot g) = \deg(1) = 0$. Hence $\deg f + \deg g = 0$. The only solution is $\deg f = \deg g = 0$. This means that f and g are "constant polynomials"$\neq 0$. Hence the claim. □

4.2 Polynomial function

Given a polynomial $f \in \mathbb{K}[X]$, $\alpha \in \mathbb{K}$, we can "evaluate" the polynomial f at α in the following way

$$f(X) = \sum_{k \in \mathbb{N}_0} a_k X^k \quad \mapsto \quad (\mathrm{eval}_\alpha f) = \sum_{k \in \mathbb{N}_0} a_k \alpha^k =: f(\alpha).$$

Now $\sum_{k \in \mathbb{N}_0} a_k \alpha^k$ is a well-defined element of $\mathbb{K}$ (note that the sum is only finite). If we vary $\alpha \in \mathbb{K}$ (but keep $f \in \mathbb{K}[X]$ fixed), we obtain a function

$$\mathbb{K} \to \mathbb{K}; \qquad \alpha \mapsto \mathrm{eval}_\alpha(f) = f(\alpha).$$

A function $\psi : \mathbb{K} \to \mathbb{K}$ is called a polynomial function if and only if there exists a polynomial $f \in \mathbb{K}[X]$ such that $\psi(\alpha) = f(\alpha)$.

Example 4.6. Let $\mathbb{K} = \mathbb{F}_2$ be the field with exactly two elements $\{\overline{0}, \overline{1}\}$. Consider $f(X) = X^2 + X$, $g(X) = 0$. Then

α	$f(\alpha)$	$g(\alpha)$
$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{0}$

Hence the two different polynomials f and g define the same polynomial function. This means we cannot identify polynomials with polynomial functions in general.

Fact: If $\mathbb{K}$ has infinitely many elements, then for $f, g \in \mathbb{K}[X]$ we have: $f = g$ if and only if the polynomial functions are the same.

Definition 4.7. $f \in \mathbb{K}[X]$, $\beta \in \mathbb{K}$ is called a zero (or a root) of f if and only if $f(\beta) = 0$.

4.3 Division with rest

In many respects the ring $\mathbb{K}[X]$ behaves similarly to the ring of integers $\mathbb{Z}$.

Proposition 4.8. *Division with rest.*
Let $f, g \in \mathbb{K}[X]$, $f, g \neq 0$, $\deg g \leq \deg f$. Then there exists unique polynomials $q, r \in \mathbb{K}[X]$ with $r = 0$ or $\deg r < \deg g$ such that

$$f = q \cdot g + r.$$

There is an easy algorithm to calculate q and r. We will show it by giving an example.

Example 4.9. $f(X) = X^3 - 1, \qquad g(X) = X + 1.$
The polynomial q is determined by multiplying g with suitable monomials to eliminate step by step the powers of f. Here

$$X^2(X+1) = X^3 + X^2 \quad \Rightarrow \quad X^2,$$
$$(X^3 - 1) - (X^3 + X^2) = -X^2 - 1.$$
$$(-X)(X+1) = -X^2 - X \quad \Rightarrow \quad -X,$$
$$(-X^2 - 1) - (-X^2 - X) = X - 1.$$
$$(1)(X+1) = X + 1 \quad \Rightarrow \quad +1,$$
$$(X - 1) - (X + 1) = -2.$$

Now we have $\deg(\text{rest}) < 1 = \deg g$. Hence the algorithm terminates and we get

$$(X^3 - 1) = \underbrace{(X^2 - X + 1)}_{q}(X + 1)\underbrace{-2}_{r}.$$

Note that it is very easy to program the algorithm for calculating the quotient polynomial and the reminder polynomial. It is also very fast if the arithmetic in the field can be implemented by a very fast algorithm.

From this result we can deduce important consequences. Given two polynomials f and g, we say f divides g if there exists a polynomial h such that $g = f \cdot h$.

Theorem 4.10. *Given $f, g \in \mathbb{K}[X]$ two polynomials. Then there exists a gcd (greatest common divisor) $d \in \mathbb{K}[X]$, $d = gcd(f, g)$ and polynomials $r, s \in \mathbb{K}[X]$ such that*

$$d = r \cdot f + s \cdot g.$$

The gcd can be calculated by the Euclid algorithm (transferred to polynomials) and the r, s by its extended version.

Recall that by definition, $d = \gcd(f, g)$ if d divides both f and g, and if k is a common divisor of f and g then it also divides d.(In fact it could also be defined as a common divisor of f and g which has maximal degree.) In fact, everything works completely as over the integers.

Warning: Already in the integer case, the gcd was only defined up to multiplication by ± 1 (the units in the ring $\mathbb{Z}$). Here we have in principal to allow it to be defined up to multiplication by the units in the ring $\mathbb{K}[X]$. The set of units is equal to $\mathbb{K}[X]^* = \mathbb{K}^*$, the non-zero scalars.

If we want to make d unique, we might require the polynomial to be normed, saying the biggest non-vanishing coefficient equals 1.

Other consequences of the division with the rest are

Proposition 4.11. *Let $f \in \mathbb{K}[X]$. Then $f(\alpha) = 0$ if and only if $(X - \alpha)$ divides f.*

Proof. Let α be a zero of f, i.e. $f(\alpha) = 0$. We show that $f(X)$ is divisible by the linear polynomial $(X - \alpha)$. For this, we apply division with rest and we get

$$f(X) = q(X)(X - \alpha) + r.$$

The polynomial f is divisible by $(X - \alpha)$ if and only if $r = 0$. Indeed, if we plug in α for X in the previous equation, we get

$$f(\alpha) = q(\alpha)(\alpha - \alpha) + r(\alpha).$$

Hence $f(\alpha) = 0 \iff r(\alpha) = 0$. But as either $(r = 0)$ or $(\deg r = 0$ and r is a constant, $r \neq 0)$, we have that $r = 0$. The opposite is obvious. □

Proposition 4.12. *Let $f \in \mathbb{K}[X]$, $\deg f = n \geq 0$. Then f has at most n zeros.*

Proof. Let α be a zero of f, i.e. $f(\alpha) = 0$. By the previous proposition we know that $f(X)$ is divisible by the linear polynomial $(X - \alpha)$. Now we set

$$f^{(1)}(X) = \frac{f(X)}{X - \alpha}.$$

This is a polynomial of degree $(n - 1)$. We check whether $f^{(1)}$ has a zero. In every step where we find a zero the degree of the quotient polynomial goes down by one. But as the degree of a polynomial is bounded by zero from below, the procedure ha to end at most after n steps. Finally we can write

$$f(X) = (X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_k) \cdot g(X),$$

with $k \leq n$ and $g(X)$ a polynomial without zeros. The zeros of f are exactly the $\alpha_1, \alpha_2, \ldots, \alpha_k$. □

Note that it is not excluded that some of the α_i coincide. In this case we get zeros of higher multiplicity. We could even say that the number of zeros counted with multiplicities is less or equal to n.

4.4 Exercises

Exercise 4.13.
Prove Proposition 4.4, i.e. prove that for $f, g \in \mathbb{K}[X]$ we have

1. $\deg(f + g) \leq \max(\deg f, \deg g)$.
2. If $\deg f \neq \deg g$, then $\deg(f + g) = \max(\deg f, \deg g)$.
3. $\deg(f \cdot g) = \deg f + \deg g$.

Exercise 4.14.
a) Calculate $\gcd(X^4 - 1, X - 1)$.
b) Calculate $\gcd(3X^3 + 2X + 1, X^2 - 4X)$.

Exercise 4.15.
a) Let p be a prime number ($p \in \mathbb{P}$). Show that

$$X^p - X = \prod_{a \in \mathbb{F}_p} (X - a).$$

b) Verify directly (over $\mathbb{F}_3$):

$$(X^3 - X) = X(X - 1)(X - 2).$$

c) Does the polynomial $X^p - X + 1$ have any zeros in $\mathbb{F}_p$?

Exercise 4.16.
Let $\mathbb{K}[X]$ be the ring of polynomials. We introduce the formal differentiation D as the map $D(X^n) = nX^{n-1}$ linearly extended to all polynomials, i.e.

$$f = \sum_{i=0}^{n} a_i X^i \quad \mapsto \quad D(f) = \sum_{i=0}^{n} i a_i X^{i-1}.$$

a) Verify

$$D(f + g) = D(f) + D(g), \quad D(\alpha f) = \alpha D(f), \alpha \in \mathbb{K},$$
$$D(f \cdot g) = D(f)g + fD(g), \quad D(g^2) = 2gD(g).$$

Is $D(X^n) = 0$ for $n \geq 1$ possible?

b) A polynomial f is called square-free if there does not exist a polynomial g of degree ≥ 1 such that g^2 divides f. Show that if $\gcd(f, D(f)) = 1$, then f is square-free.

c) Show that if α is a double zero of f, i.e. the polynomial f is divisible by $(X - \alpha)^2$, then α is also a zero of $D(f)$.

5 Field extensions

Starting from the integers $\mathbb{Z}$ we were able to construct residue class rings $\mathbb{Z}_n$ respectively fields $\mathbb{F}_p$, $p \in \mathbb{P}$, by taking the integers modulo the number n.
We already realized that the polynomial ring $\mathbb{K}[X]$ behaves algebraically in a similar manner as the integers (of course as explicit objects they are completely different). We want to extend this analogy even further by now considering "residue rings" if we consider polynomials "modulo a fixed polynomial f". In this way we obtain rings and fields $\mathbb{L}$ which contain our field $\mathbb{K}$ as subfield if the polynomial f is "irreducible" (see below).
Field extensions of the fields $\mathbb{F}_p$ are of special importance in the application in computer science.

Let $\mathbb{K}$ be a field. Recall that $(\mathbb{K}[X], +, \cdot)$ is a ring, the constant polynomial 0 is the neutral element of addition and the constant polynomial 1 the neutral element (unit) for multiplication. We fix a polynomial f and set

$$(f) := \{g \cdot f \mid g \in \mathbb{K}[X]\}$$

for the set of polynomials which are multiples of f. One verifies directly that

$$0 \in (f),$$

$$\begin{aligned} g_1, g_2 \in (f) \quad &\Rightarrow \quad g_1 + g_2 \in (f), \\ g \in (f), h \in \mathbb{K}[X] \quad &\Rightarrow \quad h \cdot g \in (f). \end{aligned}$$

Such a set is called an ideal in the ring $\mathbb{K}[X]$. But for us it is just a name for the moment.

Definition 5.1. We call two polynomials g and h equivalent (mod f), $g \sim h$, if and only if $g - h$ is divisible by f.

This can be formulated equivalently as $g - h \in (f)$, or $g = h + k \cdot f$, with $k \in \mathbb{K}[X]$.

Exercise 5.2. Show that this is an equivalence relation.

As this is an equivalence relation we can again introduce the equivalence class of g

$$\overline{g} := \{h \in \mathbb{K}[X] \mid h \sim g\} = \{h \in \mathbb{K}[X] \mid h = g + k \cdot f \text{ with } k \in \mathbb{K}[X]\},$$

and the set of equivalence classes

$$R := \mathbb{K}[X]/(f) = \{\overline{g} \mid g \in \mathbb{K}[X]\}.$$

(Everything is completely the same as in the modulo integer case.)

Definition 5.3.
$\overline{g} + \overline{h} := \overline{g+h}$,
$\overline{g} \cdot \overline{h} := \overline{g \cdot h}$,

Proposition 5.4. *$(R, +, \cdot)$ is a (commutative) ring. $\overline{0}$ is the neutral element of the addition, and $\overline{1}$ is the neutral element of the multiplication.*

Proof. We have to show that the definition of addition and multiplication of classes is well-defined. For $g, g' \in \overline{g}$, $h, h' \in \overline{h}$ we have

$$g' = g + k \cdot f, \qquad h' = h + l \cdot f \qquad \text{with } k, l \in \mathbb{K}[X].$$

Hence,

$$g' + h' = (g + h) + (k + l) \cdot f,$$
$$g' \cdot h' = (g + k \cdot f) \cdot (h + l \cdot f) = g \cdot h + (h \cdot k + g \cdot l + k \cdot l \cdot f) \cdot f.$$

In particular

$$g + h \sim g' + h', \qquad g \cdot h \sim g' \cdot h',$$

and the operations are well-defined on R.
That R is a ring follows from the fact that the operations are coming from $\mathbb{K}[X]$, and $\mathbb{K}[X]$ is a ring. □

Next we want to give a "standard" system of representing elements of the residue classes. Let n be the degree of the (fixed) polynomial f. If g is an arbitrary polynomial, we can write it in a unique way

$$g = q \cdot f + r,$$

with $r = 0$ or r is a polynomial of degree $< n$. Hence, a system of representing elements can be given as polynomials of degree $< n$, i.e.

$$g \sim \sum_{k=0}^{n-1} a_k X^k, \quad a_k \in \mathbb{K}.$$

Moreover we have the following proposition.

Proposition 5.5. *$R = \mathbb{K}[X]/(f)$ is a $\mathbb{K}$-vector space of dimension n.*

Proof. (for those who know the notion of vector spaces.)
First $\mathbb{K}[X]$ is not only a ring, but also a vector space over the field $\mathbb{K}$, the ideal (f) is a vector subspace. Hence, R as a quotient is again a vector space.
Following the remark above, every element in $\mathbb{K}[X]/(f)$ can be represented by either the zero polynomial or a polynomial of degree $< n$. In particular, the vector space will be generated by the residue classes of the monomials X^i, $i = 0, 1, \ldots, n-1$. It will be a basis if the residue classes are linearly independent. Assume that

$$a_0\overline{X^0} + a_1\overline{X^1} + a_2\overline{X^2} + \cdots + a_{n-1}\overline{X^{n-1}} = \overline{0}$$

is a linear combination of the zero element. This relation is true if and only if

$$a_0X^0 + a_1X^1 + a_2X^2 + \cdots + a_{n-1}X^{n-1} = k \cdot f$$

with a polynomial k. On the right hand side there is either the zero polynomial or a polynomial of degree $\geq n$. On the left hand side there is either the zero polynomial or a polynomial of degree $< n$. The only solution is the zero polynomial. In particular, all $a_i = 0$. Hence, the generating set is linearly independent, and as such a basis. In particular, the dimension of the vector space R over the field $\mathbb{K}$ is n. □

Proposition 5.6. *Let g be a polynomial with $gcd(g, f) = 1$. Denote by $\overline{g}$ its equivalence class. Then $\exists h \in \mathbb{K}[X]$ such that $\overline{g} \cdot \overline{h} = \overline{1}$.*

Proof. As the $\gcd(g, f) = 1$, the extended Euclid algorithm gives a presentation

$$1 = k \cdot f + l \cdot g, \qquad \text{with } k, l \in \mathbb{K}[X].$$

If we pass to the residue classes, we get

$$\overline{1} = \overline{k} \cdot \overline{f} + \overline{l} \cdot \overline{g} = \overline{k} \cdot \overline{0} + \overline{l} \cdot \overline{g} = \overline{l} \cdot \overline{g},$$

and this completes the proof. □

Warning: $\overline{g} \cdot \overline{h} = \overline{1}$ does not mean that the product of the polynomials g and h will be the polynomial which is constant to 1. In fact, for non-constant g, there will never exist such an h with $g \cdot h = 1$ by degree reasons. But $\overline{g} \cdot \overline{h} = \overline{1}$ says that $g \cdot h$ is a polynomial which is 1 plus a polynomial multiple of f.

Definition 5.7. A polynomial $f \in \mathbb{K}[X]$ is called irreducible if $\deg f \geq 1$ and f cannot be written as the product of two polynomials of degree ≥ 1. It is called reducible if it is not irreducible.

Irreducible polynomials correspond to the prime numbers in the similar case of integers.

Warning: The fact whether a polynomial is reducible or irreducible depends crucially on the base field $\mathbb{K}$ we consider. An example is the polynomial

$$f(X) = X^2 + 1,$$

which we might consider as real or complex polynomial, i.e.

$$f \in \mathbb{R}[X] \quad \text{or} \quad f \in \mathbb{C}[X].$$

It is irreducible as a real polynomial, but as a complex polynomial it decomposes as

$$f(X) = (X + i)(X - i),$$

with $i = \sqrt{-1} \in \mathbb{C}$.

Question: How about a field of characteristics 2?

Theorem 5.8. *Let $f \in \mathbb{K}[X]$ be a polynomial. Then $R := \mathbb{K}[X]/(f)$ is a field if and only if f is irreducible.*

Proof. First note that $\mathbb{K}[X]/(f)$ is a field if and only if every element different from the zero element is invertible. Note that "zero" here means zero class. And the polynomials representing the zero class are exactly those polynomials which are multiples of f.
If f is not irreducible it can be written as $f = g \cdot h$, with g, h of degree between 1 and $n - 1$. In particular, $\overline{0} = \overline{g} \cdot \overline{h}$. From the degree it follows that both $\overline{g}$ and $\overline{h}$ are not equal to $\overline{0}$. Hence, these elements are non-trivial zero-divisors. Hence, R is not a field.
Now let f be irreducible. One has to show that every non-zero element is invertible. Let $\overline{g} \neq \overline{0}$, hence g is not a multiple of f. Let $d = \gcd(f, g)$. The greatest common divisor, d has to divide f. But f is irreducible and hence, it has only divisors (up to associated elements) 1 and f itself. As d has also to divide g and g is not a multiple of f, we have that only $d = 1$ is possible. From Proposition 5.6 it follows that $\overline{g}$ is invertible. Hence R is a field. □

Having defined such an $\mathbb{L}$, we can always embed $\mathbb{K} \to \mathbb{L}$ as a subfield by assigning to $\alpha \in \mathbb{K}$ the residue class $\overline{\alpha} = \alpha \mod f$ where we consider the constant α as constant polynomial. $\mathbb{K}$ is a subfield of $\mathbb{L}$, $\mathbb{L}$ is an extension field of $\mathbb{K}$. In fact it is called an algebraic field extension as it is defined via this process.

If $\mathbb{L}$ is a field extension of $\mathbb{K}$, then $\mathbb{L}$ is always a vector space over $\mathbb{K}$. Indeed, we take the addition inside of $\mathbb{L}$ as vector space addition

$$\mathbb{L} \times \mathbb{L}, \quad (v, w) \mapsto v + w,$$

and the multiplication in $\mathbb{L}$ as multiplication with the scalars from $\mathbb{K}$

$$\mathbb{K} \times \mathbb{L}, \quad (\alpha, v) \mapsto \alpha \cdot v.$$

The necessary axioms are of course fulfilled. In general, the dimension of $\mathbb{L}$ considered as vector space could be finite or infinite. Recall that for $\mathbb{L} = \mathbb{K}[X]/(f)$ with an irreducible polynomial f its dimension over $\mathbb{K}$ equals $\deg f$ (see Proposition 5.5).

5.1 Examples of field extensions

1. As base field $\mathbb{K}$ we consider the field of rational numbers,

$$\mathbb{L} := \{\alpha + \beta \cdot \sqrt{2} \mid \alpha, \beta \in \mathbb{Q}\} \subset \mathbb{R}$$

is a subset of the real numbers. $\mathbb{L}$ is a field extension of $\mathbb{Q}$.

(a) $\mathbb{Q} \to \mathbb{L}; \quad \alpha \mapsto \alpha + 0 \cdot \sqrt{2}$.

(b) $0 \in \mathbb{L}, \quad 1 \in \mathbb{L}$.

(c) $x = \alpha + \beta \cdot \sqrt{2} \in \mathbb{L} \quad \to \quad -x = -\alpha + (-\beta) \cdot \sqrt{2} \in \mathbb{L}$. Hence $(\mathbb{L}, +)$ is a group.

(d) $x = \alpha_1 + \beta_1 \cdot \sqrt{2}, \quad y = \alpha_2 + \beta_2 \cdot \sqrt{2} \in \mathbb{L} \quad \to \quad x \cdot y = \left(\alpha_1\alpha_2 + \beta_1\beta_2\sqrt{2}^2\right) + (\alpha_1\beta_2 + \alpha_2\beta_1)\sqrt{2} = \alpha' + \beta' \cdot \sqrt{2} \in \mathbb{L}$, as $(\sqrt{2})^2 = 2 \in \mathbb{Q}$.

(e) Existence of the inverse? $x = \alpha + \beta \cdot \sqrt{2}$ (either $\alpha \neq 0$ or $\beta \neq 0$). We take

$$y = \left(\frac{\alpha}{\alpha^2 - 2\beta^2}\right) + \left(\frac{-\beta}{\alpha^2 - 2\beta^2}\right)\sqrt{2}$$

and verify that $x \cdot y = 1$.
The only problem is to exclude $\alpha^2 - 2\beta^2 = 0$. Assume that $\alpha^2 - 2\beta^2 = 0 \to \alpha^2 = 2\beta^2$. As $\alpha = \beta = 0$ was excluded, this means that in this case, both are different from zero.

$$\to 2 = \frac{\alpha^2}{\beta^2} \to \sqrt{2} = \pm\frac{\alpha}{\beta}.$$

As $\alpha \in \mathbb{Q}$ and $\beta \in \mathbb{Q}$, we have $\frac{\alpha}{\beta} \in \mathbb{Q} \to \sqrt{2} \in \mathbb{Q}$. This is a contradiction as you should have learned in school (Exercise: supply the proof).

(f)In fact this field extension can also be described as

$$\mathbb{L} = \mathbb{Q}[X]/(X^2 - 2).$$

Note that $f(X) = X^2 - 2$ is irreducible over $\mathbb{Q}$ as its zeros are $+\sqrt{2}$ and $-\sqrt{2}$ which do not lie in $\mathbb{Q}$. The element $X \mod f$ is identified with $\sqrt{2}$.

2. The set

$$\mathbb{L} = \{\alpha + \beta 2^{\frac{1}{3}} \mid \alpha, \beta \in \mathbb{Q}\} \subset \mathbb{R}$$

is not a field, as $2^{\frac{1}{3}} \cdot 2^{\frac{1}{3}} \notin \mathbb{L}$.
Exercise: verify this.
In fact the smallest subfield of $\mathbb{R}$ containing $2^{\frac{1}{3}}$ is $\mathbb{Q}[X]/(X^3-2)$. As a vector space it has a basis over $\mathbb{Q}$, for instance $1, 2^{\frac{1}{3}}, \left(2^{\frac{1}{3}}\right)^2$, where $2^{\frac{1}{3}} \Leftrightarrow X \mod (X^3-2)$, $\left(2^{\frac{1}{3}}\right)^2 \Leftrightarrow X^2 \mod (X^3-2)$.

3. Let $\mathbb{K} = \mathbb{R}$. For $i \in \mathbb{C}, i^2 = -1$,

$$\mathbb{L} = \{r + s \cdot i \mid r, s \in \mathbb{R}\}$$

is a field. This field is the field of the complex numbers $\mathbb{C}$. It can be constructed as $\mathbb{C} \cong \mathbb{R}[X]/(X^2+1)$.

5.2 Algebraic elements

Let $\mathbb{L}$ be a field extension of $\mathbb{K}$.

Definition 5.9. $\alpha \in \mathbb{L}$ is called an algebraic element over $\mathbb{K}$ if and only if $\exists f \in \mathbb{K}[X]$ (a polynomial with coefficients from $\mathbb{K}$) such that $f(\alpha) = 0$.

Remark 6.
1) If $f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ is a polynomial with coefficients in $\mathbb{K}$, then we can plug in $\alpha \in \mathbb{L}$. We get

$$f(\alpha) := a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n \in \mathbb{L}$$

if we calculate this expression in $\mathbb{L}$. Note that $a_i \in \mathbb{K} \leqslant \mathbb{L}$, $i = 0, \ldots, n$.

2) If α is algebraic over $\mathbb{K}$, then there exists also a polynomial f different from the zero polynomial, such that $f(\alpha) = 0$ and f has minimal degree.

3) Elements which are not algebraic over $\mathbb{K}$ are called transcendent over $\mathbb{K}$.

Example 5.10.
1) $i \in \mathbb{C}$ is algebraic over $\mathbb{R}$, as for $f(X) = X^2 + 1 \in \mathbb{R}[X]$, $f(i) = 0$. Indeed, i is even

algebraic over $\mathbb{Q}$ as f is a polynomial with rational coefficients.

2) π is transcendent over $\mathbb{Q}$, but algebraic over $\mathbb{R}$ as zero of the polynomial $X - \pi$.

3)All elements of $\mathbb{C}$ are algebraic over $\mathbb{R}$, but not all are algebraic over $\mathbb{Q}$.

Definition 5.11. A field $\mathbb{K}$ is called algebraically closed if every polynomial with coefficients from $\mathbb{K}$ of $\deg f \geq 1$ has at least one zero.

Consequence: All polynomials over an algebraically closed field $\mathbb{K}$ can be written as a product of linear polynomials $X - \alpha_i$, $\alpha_i \in \mathbb{K}$. The proof is an exercise (use Division Algorithm).

Example 5.12. $\mathbb{C}$ is algebraically closed.

Remark 7.
Given $\mathbb{K}$ a field, then there exists always the algebraic closure $\overline{\mathbb{K}}$ which is the minimal field extension of $\mathbb{K}$ that is algebraically closed.

Example 5.13. $\mathbb{C}$ is the algebraic closure of the field $\mathbb{R}$.

As we are especially interested in finite fields, the following is interesting.

Proposition 5.14. *Let $\mathbb{K}$ be a field, $\#K < \infty$. Then $\mathbb{K}$ is not algebraically closed.*

Proof. We consider the polynomial

$$f(X) = \prod_{a \in \mathbb{K}} (X - a).$$

This polynomial is of degree $\#\mathbb{K}$. Obviously, $f(b) = 0, \forall b \in \mathbb{K}$. Consider the polynomial $g(X) = f(X) + 1$. Then $g(b) = 1$, $\forall b \in \mathbb{K}$, but $\deg g \geq 1$. Hence $\mathbb{K}$ cannot be algebraically closed. □

6 Calculations in field extensions

Let $\mathbb{K}$ be a field in which we know how to calculate, e.g. $\mathbb{K} = \mathbb{Q}$, the rational numbers, or $\mathbb{K} = \mathbb{F}_p$, the residue class field for $p \in \mathbb{P}$ a prime number. In both fields the arithmetic is simple and can be easily done by a computer. But how to calculate in field extensions? Particularly interesting for our course is the arithmetic in the Galois fields $\mathbb{F}_{p^n}$.

Fact: If $\mathbb{L}$ is a field extension of $\mathbb{K}$ with $\dim_{\mathbb{K}} \mathbb{L} < \infty$, then $\mathbb{L}$ can be constructed by a finite sequence of field extensions:

$$\mathbb{K} = \mathbb{L}_0 \subsetneqq \mathbb{L}_1 \subsetneqq \mathbb{L}_2 \subsetneqq \cdots \subsetneqq \mathbb{L}$$

such that

$$\mathbb{L}_i = \mathbb{L}_{i-1}[X]/(f_i)$$

where f_i is a polynomial in $\mathbb{L}_{i-1}[X]$ which is irreducible over $\mathbb{L}_{i-1}$.

Hence the building blocks for field extensions are the quotients with respect to irreducible polynomials and we need to know how to calculate there.
Take $\mathbb{L} = \mathbb{K}[X]/(f)$ with f an irreducible polynomial over $\mathbb{K}$, $n = \deg f \geq 1$. We can normalize f such that it starts like $f(X) = 1 \cdot X^n + \ldots$. We write $f(X) = \sum_{i=0}^{n} a_i X^i$ with $a_i \in \mathbb{K}$ and $a_n = 1$.

A basis of the field $\mathbb{L}$ over $\mathbb{K}$ is given by the elements

$$1, \quad X \bmod f, \quad X^2 \bmod f, \quad \ldots . \quad X^{n-1} \bmod f \ .$$

We set $\alpha := X \bmod f \in \mathbb{L}$ and obtain $X^k \bmod f = (X \bmod f)^k = \alpha^k$. Hence, we can describe the basis also as $\{1 = \alpha^0, \alpha, \ldots, \alpha^{n-1}\}$.

Proposition 6.1. *Let $\mathbb{L}$ be the above field extension defined via the irreducible polynomial f, then $\alpha = X \bmod f$ is a zero of the polynomial f which lies in the extension field $\mathbb{L}$.*

Proof. We calculate

$$f(\alpha) = f(X \bmod f) = \sum_{i=0}^{n} a_i (X \bmod f)^i = \sum_{i=0}^{n} (a_i X^i \bmod f) =$$
$$(\sum_{i=0}^{n} a_i X^i) \bmod f = f \bmod f \ = \quad 0. \quad (5)$$

□

Question: Is f irreducible over $\mathbb{L}$ [1]?

Note that

$$f(\alpha) = 0 \quad \rightarrow \quad 0 = \sum_{i=0}^{n} a_i \alpha^i, \quad a_n = 1 \quad \rightarrow \quad \alpha^n = -\sum_{i=0}^{n-1} a_i \alpha^i.$$

Recall that this corresponds to

$$X^n = -\sum_{i=0}^{n-1} a_i X^i \bmod f.$$

Note that we do not need to "know" $\alpha \in \overline{\mathbb{K}}$ to be able to calculate with, as $\alpha = X \bmod f$ and we know to calculate with polynomials modulo an irreducible polynomial f.

Having the above in mind we can now discuss the arithmetic in $\mathbb{L}$

Addition: is simple, it is just addition of polynomials, given by adding the coefficients in front of the monomial.

Multiplication: multiply the two polynomials and reduce the degree of the product to a degree $< n$ by division with rest by the polynomial f.

Example 6.2. Let us consider the polynomial

$$f(X) = X^2 + X + 1$$

in parallel over the field $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{F}_2$.
First note that a degree 2 polynomial can only decompose into linear polynomials (meaning it is reducible). Hence a reducible degree 2 polynomial has to have a zero. The polynomial $f(X)$, considered as a real polynomial, has no zeros on $\mathbb{R}$ (its zeros are complex: $e^{\frac{2\pi i}{3}}$, $e^{\frac{4\pi i}{3}}$).
Also $f(X) = X^2 + X + 1 \in \mathbb{F}_2[X]$ does not have a zero as $f(1) = 1 + 1 + 1 = 1$ and $f(0) = 0 + 0 + 1 = 1$. This means

$$\mathbb{L} = \mathbb{K}[X]/(f) \cong \mathbb{K} \oplus \mathbb{K} \cdot \alpha$$

is a field extension of dimension 2, with basis the monomials in $\mathbb{K}[X]$ mod f

$$1, \quad \alpha = X \bmod f = \overline{X}.$$

[1] Of course not, as it has a zero.

It follows:

$$\alpha^2 = -\alpha - 1.$$

To illustrate the general method for multiplication, let us multiply $\alpha = \overline{X}$ and $\alpha + 1 = \overline{X} + 1$.

$$\overline{X} \cdot (\overline{X} + 1) = \overline{X}^2 + \overline{X}.$$

This is a polynomial of degree 2. Hence we have to reduce it

$$(X^2 + X) : (X^2 + X + 1) = 1, \quad \text{with rest } r = -1.$$

Hence:

$$\alpha \cdot (\alpha + 1) = -1.$$

Of course we could have directly guessed it in this simple example.

Inversion: Recall how we did inversion: given a polynomial residue class $\overline{h} \neq \overline{0}$. Then we are looking for a polynomial g such that

$$g \cdot h = q \cdot f + 1.$$

Then $\overline{g}$ is the inverse element. The method to calculate $\overline{g}$ is the extended Euclid algorithm which we discussed earlier. In our example we calculate the inverse of $\alpha = X \bmod f$. This means $\overline{h} = \alpha$ or equivalently $\overline{h} = X \bmod f$. As f is irreducible, we know that $\gcd(X, X^2 + X + 1) = 1$, but we need the algorithm to get the combination. First, $(X^2 + X + 1) = (X + 1) \cdot X + 1$. In fact we get the extension after one step. We can resubstitute:

$$1 = (X^2 + X + 1) + \underbrace{(-X - 1)}_{g} \cdot \underbrace{X}_{h}.$$

Hence

$$(\alpha)^{-1} = -\alpha - 1.$$

In this case one could for sure have guessed the result from

$$0 = \alpha^2 + \alpha + 1 \quad \rightarrow \quad 1 = -(\alpha^2 + \alpha) = \alpha \cdot (-\alpha - 1).$$

In fact, in general for α, the root of the irreducible polynomial f, the α^{-1} can always be directly calculated. Let

$$f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0$$

be the irreducible polynomial for α. Then $a_0 \neq 0$ (otherwise f would have a factor of X and would not be irreducible). From $f(\alpha) = 0$ we can obtain

$$-a_0 = \alpha \cdot (\alpha^{n-1} + a_{n-1}\alpha^{n-2} + \cdots + a_1)$$

or

$$1 = \alpha \cdot \underbrace{\left(\frac{1}{-a_0} \cdot (\alpha^{n-1} + a_{n-1}\alpha^{n-2} + \cdots + a_1) \right)}_{\alpha^{-1}}.$$

All calculations were done without knowing α, we only needed to know that f is the irreducible minimal polynomial for α.

Remark 8. Also there was nowhere a specific reference in the example whether our field was $\mathbb{K}$ or $\mathbb{F}_2$.

Remark 9. Indeed we could give another description of our field $\mathbb{L}$ as follows. Our polynomial f over $\overline{\mathbb{K}}$ (the algebraic closure) will have n zeros, α_i, $i = 1, \ldots, n$. Let α_i one of these zeros, then

$$\mathbb{L} = \mathbb{K}[X]/(f) \cong \mathbb{K}[\alpha_i] \geqslant \overline{\mathbb{K}}. \tag{6}$$

Here $\mathbb{K}[\alpha_i] := \{g(\alpha_i) \mid g \in \mathbb{K}[X]\}$. These are all polynomial combinations in α_i. Note that as α_i is a zero of f we have $f(\alpha_i) = 0$. Given an arbitrary polynomial g we can make a division by f with rest:

$$g(X) = q(X) \cdot f(X) + r(X)$$

with $r \equiv 0$ or $0 \leq \deg r < \deg f$. If we plug in α_i we get:

$$g(\alpha_i) = q(\alpha_i) \cdot f(\alpha_i) + r(\alpha_i) = 0 + r(\alpha_i).$$

Hence only those $g \in \mathbb{K}[X]$ are needed of degree $< \deg f = n$. In the above identification (6), the correspondence is as follows:

$$\begin{aligned}
1 \quad &\mapsto \quad \alpha_i^0 = 1, \\
X \bmod f \quad &\mapsto \quad \alpha_i^1, \\
X^2 \bmod f \quad &\mapsto \quad \alpha_i^2, \\
&\vdots \\
X^{n-1} \bmod f \quad &\mapsto \quad \alpha_i^{n-1}, \\
X^n \bmod f \quad &\mapsto \quad \alpha_i^n.
\end{aligned}$$

In fact the last line is not needed anymore as $X^n \bmod f$ respectively α_i^n can be already expressed by the other terms via the polynomial f. The identification is compatible with our field operations. Hence, it is a field isomorphism. Here a warning is in order. Let α_i and α_j be two distinct zeros. We obtain from (6) that

$$\mathbb{K}[\alpha_i] \cong \mathbb{L} \cong \mathbb{K}[\alpha_j.$$

This says that as abstract fields the construction does not depend on the choice of the zero. In other words the isomorphy type of the finite fields $\mathbb{K}[\alpha_k]$ is fixed. But in general, it will not necessarily be the case that as subfields of $\overline{\mathbb{K}}$ the fields $\mathbb{K}[\alpha_i]$ and $\mathbb{K}[\alpha_j]$ coincide if $i \neq j$.

Exercise 6.3.
Consider the example (6.2) for $\mathbb{K} = \mathbb{F}_2$. In this case we only have the elements $\{0, 1, \alpha, \alpha + 1\}$ in $\mathbb{F}_2[X]/(X^2 + X + 1) = \mathbb{L}$. Write down the addition table and the multiplication table with our knowledge now (assuming the statement over the finite fields, $\mathbb{L}$ is the Galois field $\mathbb{F}_{2^2}$). We will return to this in the next section.

Exercise 6.4. Let $f(X) = X^2 - 2$ be defined over $\mathbb{Q}$. Show that it is irreducible. Describe $\mathbb{L} = \mathbb{Q}[X]/(f)$. Consider also the construction explained by Remark 9. Decompose the polynomial f into irreducible factors over $\mathbb{L}$.

Exercise 6.5. Do the same for $f(X) = X^3 - 2$ (defined over $\mathbb{Q}$). Show that it is irreducible. Describe $\mathbb{L} = \mathbb{Q}[X]/(f)$. Consider also the construction explained by Remark 9. Decompose the polynomial f into irreducible factors over $\mathbb{L}$.

7 Finite fields

In this section, we concentrate on finite fields. Beside their mathematical interest these fields also play an important role in applications in Computer Sciences. In particular, they show up in cryptography and coding theory.
We already quoted the first part of the following theorem:

Theorem 7.1. *1. Let $\mathbb{K}$ be a field with finitely many elements, then there exists a prime number $p \in \mathbb{P}$ and $n \in \mathbb{N}$ such that $\#\mathbb{K} = q = p^n$, and $\mathbb{K} \cong \mathbb{F}_q$, where $\mathbb{F}_q$ is the field consisting of zeros of the polynomial $f(X) = X^q - X$, $f \in \mathbb{F}_p[X]$.*

2. For every $p \in \mathbb{P}$, $n \in \mathbb{N}$, there exists a finite field $\mathbb{F}_q$ which has exactly $q = p^n$ elements.

In the following we will show this theorem and give additional information about their structure. We remark that this is not only of theoretical interest, as we, by the constructions, will understand how to calculate in these fields. This is needed in applications. Recall that $\mathbb{F}_p = \mathbb{Z}_p$ ($p \in \mathbb{P}$) is a field which has exactly p elements. Its elements are given by

$$\{\overline{0}, \overline{1}, \ldots, \overline{p-1}\}$$

where $\overline{a} = a \mod p$. Recall also that the characteristics char $\mathbb{K}$ of a field $\mathbb{K}$ with unit element 1 is defined to be the smallest element $n \in \mathbb{N}$ with $n \cdot 1 = \underbrace{1 + \cdots + 1}_{\text{n-times}} = 0$ (if such an n exists, otherwise char $\mathbb{K} = 0$). We have char $\mathbb{F}_p = p$.

Proposition 7.2. *Let $\mathbb{K}$ be a finite field. Then $\exists p \in \mathbb{P}$ such that char $\mathbb{K} = p$ and $\mathbb{K}$ is a field extension of $\mathbb{F}_p$ and its dimension as vector space over $\mathbb{F}_p$ is finite, i.e.* $\dim_{\mathbb{F}_p} \mathbb{K} = n$. *Moreover $\#K = p^n$.*

Proof. (of the proposition)
Let 1 be the (multiplicative) unit of $\mathbb{K}$. First as $\#K = q$ is finite we have that the set $\{n \cdot 1 \in \mathbb{K} \mid n \in \mathbb{N}\}$ has only finitely many different elements. Hence there must be $n_1, n_2 \in \mathbb{N}$ (without restrictions) such that $n_2 > n_1$ with $n_1 \cdot 1 = n_2 \cdot 1$. This implies that $(n_2 - n_1) \cdot 1 = 0$. Hence, char $\mathbb{K} \neq 0$. But we showed earlier that in this case char $\mathbb{K}$ must be a prime number p. We fix this prime number. Inside the field $\mathbb{K}$ we have a subfield $\mathbb{L} = \{n \cdot 1 \mid n = 0, 1, \ldots, p-1\}$ which is isomorphic to $\mathbb{F}_p$ via the map $1 \mapsto \overline{1} = 1 \mod p$ (you have to verify this as an exercise). We identify $\mathbb{L}$ with $\mathbb{F}_p$. Now $\mathbb{K}$ is a field extension over $\mathbb{F}_p$. But as explained before, field extensions over a base field are always vector spaces over the base field. As $\mathbb{K}$ has only finitely many elements, the dimension of $\mathbb{K}$ over $\mathbb{F}_p$ will be finite. Let us denote this dimension by n. By the very

definition of dimension we can find n basis elements $v_1, \ldots, v_n \in \mathbb{K}$ such that every element $u \in \mathbb{K}$ can be given in a unique way as a linear combination

$$u = \sum_{i=1}^{n} \alpha_i v_i, \qquad \alpha_i \in \mathbb{F}_p.$$

Vice versa every such linear combination is an element of $\mathbb{K}$. For every coefficient α_i we have p different choices. This makes $\underbrace{p \cdot p \cdot \cdots \cdot p}_{\text{n-times}} = p^n$ different elements. Hence the claim $\#K = p^n$. □

Next we consider the multiplicative group $(\mathbb{K}^* = \mathbb{K} \setminus \{0\}, \cdot)$. As $\#\mathbb{K}^* = q - 1$ $(q = p^n)$, Little Fermat says that $\forall \alpha \in \mathbb{K}^*$:

$$\alpha^{q-1} = 1.$$

If we multiply this with α we get $\alpha^q - \alpha = 0$ which is also true for $\alpha = 0$. Hence we get that $\alpha \in \mathbb{K}$ implies that α is a zero of the polynomial $f(X) = X^q - X \in \mathbb{F}_p[X]$.
How many zeros does the polynomial f have? First $\deg f = q$. Therefore f has at most q zeros. But as $\mathbb{K}$ as q elements we get all zeros, and all are pairwise different. Hence the two sets coincide.
In the language of algebra, $\mathbb{K}$ is the splitting field of the polynomial $X^q - X$. One shows there (essentially using finer analysis of the construction of field extensions presented in the last section) that two such splitting fields are isomorphic. Hence two such $\mathbb{K}$ and $\mathbb{K}'$, with $\#\mathbb{K} = \#\mathbb{K}'$, are isomorphic.

What is missing is the <u>existence</u> of such a field extension.
First we start with $\mathbb{F}_p$ and consider the polynomial $f(X) = X^q - X$, which has q roots. As $f(0) = 0$ (and $f(1) = 0$) it will not be irreducible. But we can decompose it into a product of irreducible factors g_i $(\deg g_i > 1)$

$$f(X) = X \cdot (X - \overline{1}) g_1(X) \ldots g_r(X).$$

We take g_1, then construct the field extension

$$\mathbb{L}_1 = \mathbb{F}_p[X]/(g_1).$$

Note that this also shows us how to calculate in $\mathbb{L}_1$. Now we decompose f in $\mathbb{L}_1[X]$, and take again an irreducible factor etc. This process terminates if $f(X)$ decomposes completely into linear polynomials. But this means that all zeros of f lie in the last constructed field $\mathbb{L}$. And indeed the field $\mathbb{L}$ is the splitting field.

The problem is that a priori $\mathbb{L}$ could be much bigger than just the set of all zeros. For example, if α, β are zeros then why should $\alpha + \beta$ be also a zero?
There is an easy counterexample. Consider the polynomial $f(X) = X^2 - 1$ over $\mathbb{Q}$. Its zeros are $\sqrt{2}$ and $-\sqrt{2}$. Now $+\sqrt{2} - (-\sqrt{2}) = 2\sqrt{2}$ is obviously not a zero of f.

That in our case $\alpha + \beta$ is a zero has to do with the peculiar form of the polynomial and that char $\mathbb{F}_p = p$.

We set $E := \{\alpha \in \mathbb{L} \mid f(\alpha) = 0\}$ (for $f(X) = X^q - X$) and we will show that E is a field. Then it is clear that $\#E = q = p^r$, E is the splitting field of the polynomial and $E \cong \mathbb{L}$.

Lemma 7.3. *For $a, b \in \mathbb{L}$, $\mathbb{F}_p \leqslant \mathbb{L}$ (char $\mathbb{L} = p$), we have*

$$(a + b)^p = a^p + b^p.$$

Proof. By the binomial formula we get

$$(a + b)^p = a^p + \sum_{k=1}^{p-1} \binom{p}{k} a^{p-k} b^k + b^p,$$

where

$$\binom{p}{k} = \frac{p!}{k!(p-k)!} \in \mathbb{N}.$$

As p is prime and $k < p$, p does not divide any number between 1 and k. Therefore $p \nmid k!$ and the same is true for $p \nmid (p-k)!$. Hence $\binom{p}{k} = \frac{p \cdot (p-1)!}{k!(p-k)!} = p \cdot m$, $m \in \mathbb{N}$. As char $\mathbb{L} = p$ this means that $\binom{p}{k} 1 = 0$ in $\mathbb{L}$, the middle terms in the sum vanish and we get the result. □

Now we take $q = p^r$ and we get

$$(a + b)^q = (a + b)^{p^r} = ((a + b)^p)^{p^{r-1}} = (a^p + b^p)^{p^{r-1}} = \cdots = a^q + b^q.$$

Similarly

$$(a - b)^q = ((a)^q + (-b)^q) = a^q - b^q.$$

In the last step we used $(-1)^q = -1$ for p odd and $-1 = +1$ for $p = 2$.

Proposition 7.4. *Let $a, b \in E$. Then $a + b$, $a - b$, $a \cdot b$ and $\frac{a}{b}$ (for $b \neq 0$) are also in E. Hence E is a subfield of $\mathbb{L}$.*

Proof. $a \in E \quad \to \quad a^q - a = 0 \quad \to \quad a^q = a$. Hence

$$(a \pm b)^q = a^q \pm b^q = a \pm b \quad \to \quad a \pm b \in E,$$

$$(a \cdot b)^q = a^q \cdot b^q = a \cdot b \quad \to \quad a \cdot b \in E,$$

$$\left(\frac{a}{b}\right)^q = \frac{a^q}{b^q} = \frac{a}{b} \quad \to \quad \frac{a}{b} \in E.$$

□

This finally shows Theorem 7.1 about existence and uniqueness of the finite fields.

Theorem 7.5. *$\mathbb{F}_q^*$, the multiplication group of $\mathbb{F}_q$, is a cyclic group of order $q - 1 = p^r - 1$.*

The proof goes via showing that the zeros of $X^{q-1} - 1$ constitute a cyclic group. We will not give the details.

This theorem has important consequences. It was explained in the section on cyclic groups that after the choice of a generator the group operations are easily done via integer calculations with the exponents. Hence, after determining a generator of the multiplicative group the multiplication in the field can be done easily.

Example 7.6. Let us consider the field which has exactly 4 elements. First as $4 = 2^2$ there exists such a field. It has characteristics 2 and is the set of zeros of the polynomial

$$X^4 - X = X(X^3 - 1) = X(X - 1)(X^2 + X + 1).$$

Take $g(X) = X^2 + X + 1$. As $g(\overline{0}) = \overline{1}$, $g(\overline{1}) = \overline{1}$. g is irreducible and we can use our construction method of the last section. Set $\alpha = X \mod X^2 + X + 1$. Then

$$\alpha^2 = -\alpha - 1 = \alpha + 1 \qquad (\text{char } \mathbb{L} = 2).$$

In this case the polynomial $X^4 - X$ already completely decomposes after the first step and we get

$$\mathbb{F}_4 \cong \mathbb{F}_2[X]/(X^2 + X + 1).$$

If we write down the addition and multiplication table we obtain

$+$	0	1	α	$1+\alpha$
0	0	1	α	$1+\alpha$
1	1	0	$1+\alpha$	α
α	α	$1+\alpha$	0	1
$1+\alpha$	$1+\alpha$	α	1	0

$\cdot$	0	1	α	$1+\alpha$
0	0	0	0	0
1	0	1	α	$1+\alpha$
α	0	α	$1+\alpha$	1
$1+\alpha$	0	$1+\alpha$	1	α

Question 1: Is $(\mathbb{F}_4, +)$ a cyclic group?

Answer: No, as in the counter diagonal we get always the 0 (neutral element for the addition). All elements have order ≤ 2. But for being a cyclic group, we would need an element of order 4.

In fact: $(\mathbb{F}_{p^n}, +)$ with $n > 1$ is never a cyclic group. As the additive group is equal to the vector space additive group structure we get

$$(\mathbb{F}_{p^n}, +) \cong \underbrace{C_p \times C_p \times \cdots \times C_p}_{\text{n-times}},$$

where C_p is the cyclic group of order p. In particular, $(\mathbb{F}_{p^n}, +) \ncong \mathbb{Z}/p^n\mathbb{Z}$ for $n > 1$ as the latter is always cyclic.

Question 2: Is $(\mathbb{F}_4^*, \cdot)$ a cyclic group?

Answer: Yes. Of course this is always true by the general result. But in this case we even find a simpler argument: $\#(\mathbb{F}_4^*) = 4 - 1 = 3$, which is a prime number, hence the group is cyclic. A generator is the element α.

Remark 10. Finite fields and fields of characteristics 0 are examples of *perfect fields*. We will not give the mathematical definition of a perfect field here. For us important is the mathematical result, that for a finite-dimensional field extension $\mathbb{L}$ over a perfect field $\mathbb{K}$ there exists always an irreducible polynomial $g \in \mathbb{K}[X]$ such that $\mathbb{L} = \mathbb{K}[X]/(g)$. Necessarily, the polynomial g has to have the degree given by the dimension of the field extension.

Example 7.7. Next we consider the field which has exactly 8 elements. As $8 = 2^3$ there exists exactly a field $\mathbb{F}_8$ of this type (up to isomorphy). It is a field extension of $\mathbb{F}_2$ of dimension 3. Hence, in view of the above remark and the theorem about uniqueness (up to isomorphy) of this field it is enough to find a polynomial of degree 3 which is irreducible. A degree 3 reducible polynomial has to have as divisor a linear polynomial, hence a zero. Consequently, we search degree 3 polynomials without zeros in $\mathbb{F}_2$. We start with the ansatz $h(X) = X^3 + \alpha X^2 + \beta X + \bar{1}$, with $\alpha, \beta \in \{\bar{0}, \bar{1}\}$. We obtain

$h(\bar{0}) = \bar{1}$ and $h(\bar{1}) = \alpha + \beta$. Hence, $h(\bar{1}) \neq 0$ gives two solutions $\alpha = \bar{0}, \beta = \bar{1}$ and $\alpha = \bar{1}, \beta = \bar{0}$. Consequently we have exactly two irreducible polynomials of degree 3:

$$h_1(X) = X^3 + X^2 + 1, \qquad h_2(X) = X^3 + X + 1 \ .$$

It does not matter which one we take. The obtained fields are isomorphic (due to our general theorem about finite fields). For example, we have $\mathbb{F}_8 = \mathbb{F}_2[X]/(h_1)$. Now h_2 also decomposes completely in this field. Note that $h_2(X) = h_1(X + 1)$.

Recall that $\mathbb{F}_8$ is also the splitting field of the polynomial $F(X) = X^8 - X$. In a first step we can write over $\mathbb{F}_2$ this polynomial as

$$F(X) = X(X - 1)(X^6 + X^5 + X^4 + X^3 + X^2 + X + 1).$$

The complementary factor cannot be irreducible as otherwise the field extension has to have degree at least 6 and not 3. A direct calculation shows that the complementary factor is the product of h_1 and h_2,

$$F(X) = X(X - 1)(X^3 + X^2 + 1)(X^3 + X + 1).$$

Example 7.8. Next we consider $\mathbb{F}_9 = \mathbb{F}_{3^2}$. Hence, it is a field extension of the residue class field $\mathbb{F}_3$ of dimension 2. Consequently we have to look for a quadratic polynomial without a zero in $\mathbb{F}_3$. A general ansatz and checking all elements of the field for being a zero gives the following irreducible polynomials:

$$X^2 + 1, \quad X^2 + X + 2, \quad X^2 + 2X + 2.$$

Hence, we could again take an arbitrary one of those and get e.g. $\mathbb{F}_9 = \mathbb{F}_3[X]/(X^2 + 1)$.

Our general polynomial decomposes over $\mathbb{F}_3$ as

$$F(X) = X^9 - X = X(X - 1)(X - 2)(X^2 + 1, X^2 + X + 2)(X^2 + 2X + 2).$$

Remark 11. In the examples above the additional polynomials appearing in $X^q - X$ were always of the degree given by the field extension. There is a warning in order. The complete result is the following (can be shown by some more mathematics)

Proposition 7.9. *Let $g(X) \in \mathbb{F}_p[X]$ be a normalized (meaning the highest coefficient is 1) irreducible polynomial of degree d . Then $g(X)$ divides $X^{p^n} - X$ if and only if d divides n.*

In the above examples the only divisors of n were 1 and n itself. But for $n = 4$ this changes. In this case we get for the splitting

$$X^{16} - X = X(X - 1)(X^2 + X + 1)(X^4 + X + 1)(X^4 + X^3 + 1)(X^4 + X^3 + X^2 + X + 1) \ .$$

In particular there is also the irreducible polynomial of degree two. To construct $\mathbb{F}_{16}$ we can take any of the degree 4 factors and get e.g.

$$\mathbb{F}_{16} = \mathbb{F}_2[X]/(X^4 + X + 1) \ .$$

Recall that in $\mathbb{F}_{16}$ all factors of $X^{16} - X$ completely decompose into linear polynomials. The appearance of the degree two polynomial corresponds to the fact, that $\mathbb{F}_{2^2}$ is a subfield of $\mathbb{F}_{2^4}$.

Remark 12. Related to the proof of Theorem 7.5 is the fact that at least one of the irreducible polynomials g_i in the decomposition of $X^{p^n} - X$ to be taken to construct the field $\mathbb{F}_{p^n}$ can be chosen so that $X \bmod g_i$ will be a generator of the multiplication group of the field.

A Appendix: Public key encryption – RSA

This appendix can be studied after Section 3.
As an application of some of the techniques of residue class ring calculations we explain the RSA-crypto system (RSA = Rivest-Shamir-Adleman). It is based on the fact that at the moment there are no feasible algorithms to factorize a huge number into its prime factors, but there are reasonable algorithms to check whether a number is prime. Hence given two numbers p and q from which we know that they are prime (we checked this) we calculate the product $N = p \cdot q$. Even if other people get the number N they might be able to check that it is not prime, but will not be able to find p and q in reasonable time. In the following we will do arithmetic in the residue ring $\mathbb{Z}_N$ (which is not a field), with $N = p \cdot q$, $p, q \in \mathbb{P}$. Recall

Proposition A.1. $\#(\mathbb{Z}_N)^* = \varphi(N) = \varphi(p \cdot q) = (p-1)(q-1)$.

Proof. We have to count the numbers of $a \in \mathbb{N}$, $0 < a \leq p \cdot q$, which are relatively prime to p and q (i.e. to N). This equals the number $\varphi(N)$. One has $p \cdot q$ possible choices for a.
$\{1 \cdot p, 2 \cdot p, \ldots, p \cdot q\}$ exactly are divisible by p.
$\{1 \cdot q, 2 \cdot q, \ldots, p \cdot q\}$ exactly are divisible by q.
Only $p \cdot q$ is counted twice. Hence we get

$$p \cdot q - q - p + 1 = (p-1)(q-1)$$

which are relatively prime to p and q. □

Lemma A.2. *$M \in \mathbb{N}$, $l \in \mathbb{Z}$. Then*

$$M^{1+l(p-1)(q-1)} \equiv M \mod N.$$

Proof. We start by the remark that

$$M^{1+l(p-1)(q-1)} \equiv M \cdot \left(M^{\varphi(N)}\right)^l \equiv M \mod N$$

for $\gcd(M, N) = 1$ by the Little Fermat. Hence it is true for them.

But in fact it is also true for all the others. Little Fermat now for mod p (or mod q) says $M^{p-1} \equiv 1 \mod p$ for $\gcd(M, p) = 1$. Moreover we have for $u \in \mathbb{Z}$ that

$$M^{u \cdot (p-1)} \equiv (M^{p-1})^u \equiv 1^u \equiv 1 \mod p.$$

If we multiply this by M we get

$$M^{1+u \cdot (p-1)} \equiv M \mod p \qquad (u \in \mathbb{Z}),$$

which is true for all M.

In the same way we obtain

$$M^{1+v\cdot(q-1)} \equiv M \mod q \qquad (v \in \mathbb{Z}).$$

In particular

$$M^{1+ul(p-1)(q-1)} \equiv M \mod p \quad \textbf{and} \quad \mod q.$$

This means $M^{1+l(p-1)(q-1)} - M$ is both divisible by p and by q. Hence by the product $N = p \cdot q$, which shows the claim. Here we use that p and q are distinct primes. □

Next we describe the steps in the crypto-scheme:

1. The scheme is non-symmetric. There is A (usually called Alice) who sets up the scheme and she wants to get from B (called Bob) secret information without the problem of handing over a secret key to B.

2. First the information has to be transfered into numerical data, which has to be chopped into pieces of certain sizes. Let us assume for example the unit size to be 1000 bits. Then the information piece M is given by a number M, $0 \leq M < 2^{1000}$ (1000 bits correspond to 125 characters of 1 byte).

3. Alice selects two extremely large primes p and q such that

   $$N = p \cdot q > 2^{1000}.$$

   Next she chooses e such that $\gcd(e, p-1) = \gcd(e, q-1) = 1$. e can be chosen by trial and error. Note that calculating the gcd is very effective. The pair (N, e) will be the public key.

4. Alice calculates her secret deciphering key d. It is defined by the property

   $$d \cdot e \equiv 1 \mod (p-1)(q-1).$$

   Note that because e does not have any common prime factor neither with $(p-1)$ nor with $(q-1)$, it does not have any with their product $(p-1)(q-1)$. Hence $1 = \gcd(e, (p-1)(q-1))$ and the extended Euclid Algorithm yields d (in very fast time).

5. Now she sends out (N, e) to Bob who wants to send Alice a message M. Bob calculates

   $$C = M^e \mod N,$$

   which is easily done, and sends C to Alice.

6. Alice receives C. She knows the d associated to e and with $e \cdot d = 1 + u(p-1)(q-1)$. Hence
$$C^d = (M^e)^d \equiv M^{e \cdot d} \equiv M^{1+u(p-1)(q-1)} \equiv M \mod N$$
by our lemma. Hence she can recover the message M.

7. Everybody who wants to decipher the message C needs to know d. The knowledge of e is not enough for it. For this he has to know the number $A = (p-1)(q-1)$, then it would be an easy task. But knowing $(p-1)(q-1)$ is equivalent to knowing the factorization $N = p \cdot q$.

   Proof. a) Of course if one knows the factorization of N into $p \cdot q$, one knows $(p-1)(q-1)$.
   b) Let us assume that the number $A = (p-1)(q-1)$ is known. We calculated $(p-1)(q-1) = p \cdot q - (p+q) + 1$. We set $B = p+q$ and we see that $B = N - A + 1$, hence we know B, too. Now consider the polynomial
   $$(X-p)(X-q) = X^2 - (p+q)X + N = X^2 - B \cdot X + N.$$
   This polynomial is a quadratic polynomial with zeros p and q, which can be easily calculated. Hence knowing $A = (p-1)(q-1)$ is as hard as knowing the factorization. □

Of course there are variants also for calculating other cryptography scenarios.

Identity Check
B wants to check whether A is still the same who had sent out (N, e) to him. He sends a message a to A. Now A calculates $b = a^d$ and sends it to B. Now B calculates
$$b^e = (a^d)^e \equiv a \mod N.$$
Only A could produce the correct b.

Electronic signature
A gives out as usual the public key (N, e) and keeps her secret key d. She wants to send a message c to B. She sends $(c, c^d) = (c, f)$. Now B checks whether $f^e = c$. If so he accepts that this comes from A. Of course the message (c, c^d) can additionally be encoded also, in order to make the content c secret, too.

B Appendix: Public key encryption – DLL

The discrete logarithm problem (DLP). Let G be a finite group with neutral element e. Let a be an element of G and $\langle a\rangle$ the cyclic subgroup of G generated by a, i.e.

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

Of course $\langle a\rangle$ will only have finitely many different elements. We know $\mathrm{ord}(a) = \#\langle a\rangle$ hence

$$\langle a\rangle \ := \{a^0 = e, a^1, a^2, .., a^{\mathrm{ord}(a)-1}\}\,.$$

Recall that the group G itself is called a *cyclic group* if an $a \in G$ exists with $G = \langle a\rangle$. Such an a is called a *generator* of G. Given $a \in G$ and $k \in \mathbb{N}$ then calculating $b = a^k$ is simple (assuming that we have a group in which we can easily do multiplication).

But: Given b and a searching k such that $a^k = b$ is typically difficult. Finding such an k is called the *discrete logarithm problem (DLP)* as "$k = \log_a b$". One "brute force" approach is to calculate all a^k one after each other for all k and individually compare the result with b. This is not feasible if our group G is large.

Before discussing suitable groups I like to show how this can be used to exchange a secret key c between partners A and B. The procedure is called *Diffie-Helmann key exchange*.

1.The parties A and B choose a cyclic group G and a generator a, i.e. the pair (G, a). These data can be revealed to everybody.

2. Now A chooses a secret key k and calculates a^k, and B chooses a secret key l and calculate a^l.

3. They exchange the calculated values.

4. Hence A receives a^l, takes its kth power and calculates $(a^l)^k = a^{l\cdot k}$. The party B receives a^k, takes its lth power and calculates $(a^k)^l = a^{l\cdot k}$.

5. The common value $c = a^{l\cdot k}$ is their shared secret.

6. Known (as it went over the transmission channel) are a, a^l, a^k but neither k, nor l nor $c = a^{l\cdot k}$. To determine c one would have to know l or k. In other words, one would need to calculate the discrete logarithm of a^l or a^k.

Encoding (this means calculating a^k) can be done very effectively by squaring and multiplying as the Figure 1 shows.

This gives $O(\log k)$ operations.

Whether decoding is simple or not depends on the realization of the cyclic group.

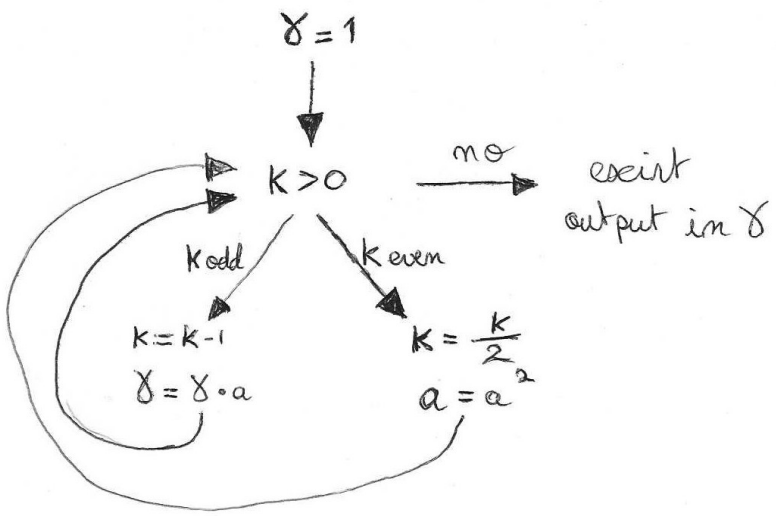

Figure 1: operation diagram

1. The group $(\mathbb{Z}_n, +)$ with generator $a = 1$. In this case the discrete logarithm problem is totally trivial. Given "a^l" means that we have l directly given. The group operations are residue class additions.

2. The multiplicative group $\mathbb{F}_q^*$ is also a cyclic group of order $(q-1)$ where $q = p^n$, $p \in \mathbb{P}$. It is more secure but q must be quite large. The group operations are polynomial multiplication combined with residue class multiplication.

3. Much better is the group $E(\mathbb{F}_q)$ of $\mathbb{F}_q$ valued points on an elliptic curve defined over a finite field (suitable chosen). In general it will not be a cyclic group, but we take cyclic subgroups of it. The group operations are still reasonable simple but the discrete logarithm problem is hard. It is a very effective method.

   In a little bit more details: An elliptic curve in the projective plane over a field $\mathbb{K}$ is given by a homogeneous cubic polynomial. The points on the curve are the points in the plane for which the defining polynomial evaluated will give zero. To be an elliptic curve, not just a cubic one, we require that there are no "singularities". One can introduce a commutative group structure $\oplus$ on the set of solutions E. Of special importance in our context is the case of $\mathbb{K}$ being a finite field $\mathbb{F}_{p^n}$. The number of points in $\mathbb{P}^2(\mathbb{F}_q)$ is finite. In fact, it is equal to q^2+q+1. In particular, $(E, \oplus)$ is a finite group. In general it will not be a cyclic group. But if we fix a point $Q \in E$ we obtain the finite cyclic subgroup $\langle Q \rangle$ generated by Q. In fact with a suitable Q we obtain cyclic groups which have many elements. Calculations in E (and hence in $\langle Q \rangle$) are rather easy. Nevertheless, the discrete logarithm problem is very hard to solve - at least if some special curves are excluded. Such groups are needed in cryptography. The method is now widely used.

As an example used in the context of bitcoins [2] we quote the following. One takes F_p with p the prime number

$$p = 2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1, \tag{7}$$

which is slightly less than 2^{256}. One considers the elliptic curve E given by the polynomial

$$f(X, Y, Z) = Y^2Z - X^3 - 7Z^3. \tag{8}$$

As generator for the cyclic subgroup considered one takes the point $Q = (q_x : q_y : 1)) \in E$ with coordinates[3]

$q_x = 55066263022277343669578718895168534326250603453777594175500187360389116729240,$

$q_y = 32670510020758816978083085130507043184471273380659243275938904335757337482424.$

For more details on the fascinating applications of elliptic curves in crypto, see the more specialized literature.

[2] Ricardo Pérez-Marco: *Bitcoins and decentralised trust protocols*, Newsletter of the EMS, Vol. 100, 31–38 (2016)

[3] Up to copy-errors.

C Appendix:Equivalence relations, coset spaces and quotient groups

C.1 Some remarks on equivalence relations

We start by discussing in a more general context equivalence relations. We used them already twice in the course and they show up everywhere.

Let M be a set. A relation R is a subset of $M \times M$ ($R \subset M \times M$). We denote xRy is true if $(x, y) \in R$ (in words: x is in relation with y). We say xRy is false if $(x, y) \notin R$.

For $M = \mathbb{Z}$ we have a lot of relations, e.g. x is in relation to y if
(1) $x \leq y$,
(2) $x = y$,
(3) $x - y$ is divisible by two,
(4) $x < y$,
and so on.

A relation is a generalization of a function. If for instance $f : \mathbb{R} \to \mathbb{R}$ is a function, we can introduce in $\mathbb{R} \times \mathbb{R}$ the relation $xRy \leftrightarrow y = f(x)$. For functions we have the special structure that $\forall x \in \mathbb{R}$ we have just one element $y \in \mathbb{R}$ such that xRy. Be warned: this is not symmetric. If a function is not injective, there might be $x_1, x_2 \in \mathbb{R}$, $x_1 \neq x_2$ but x_1Ry and x_2Ry as $y = f(x_1) = f(x_2)$ is not excluded.

Relations are not only restricted to numbers. Let R be a relation given on a set M. We call the relation R

1. reflexive $\leftrightarrow$ $\forall a \in M$: aRa.
2. symmetric $\leftrightarrow$ $(aRb \to bRa)$.
3. transitive $\leftrightarrow$ (aRb and $bRc \to aRc$).

Definition C.1. A relation R on a set M is called an equivalence relation if R is reflexive, symmetric, and transitive.

Usually an equivalence relation R is denoted by $\sim$, i.e. $x \sim y$. Examples are of course the equality of numbers (or elements of a general set) and the equivalence mod N discussed in Section 3.

$$a, b \in \mathbb{Z}, \quad a \sim b \leftrightarrow b - a \text{ is divisible by } N.$$

Given M a set and an equivalence relation $\sim$ we introduce

$$\overline{a} := \{b \in M \mid b \sim a\} \subset M.$$

$\overline{a}$ is a subset of M, as $a \sim a$, $a \in \overline{a}$, it is non-empty.

Proposition C.2. $\overline{a} = \overline{b} \quad \leftrightarrow \quad b \sim a.$

Proof. If $\overline{a} = \overline{b}$, then $b \in \overline{a}$ (as $b \in \overline{b}$). Hence $b \sim a$.
Vice versa, $\forall c \in \overline{b}$, we have $c \sim b$ if $b \sim a$. By transitivity we get $c \sim a$. This implies $c \in \overline{a} \to \overline{b} \subset \overline{a}$. Interchanging the role of $\overline{a}$ and $\overline{b}$ we get $\overline{b} = \overline{a}$. □

Proposition C.3.

$$\overline{a} \cap \overline{b} = \begin{cases} \varnothing & \text{if } a \not\sim b, \\ \overline{a} & \text{if } a \sim b. \end{cases}$$

Proof. $a \not\sim b$ is short for not $(a \sim b)$.
If $c \in \overline{a} \cap \overline{b}$, then $c \sim a$ and $c \sim b$. By symmetry, $a \sim c$ and $c \sim b$. So by transitivity we get $a \sim b$. Hence if there is an element in $\overline{a} \cap \overline{b}$, $a \sim b$ and this says that $\overline{a} = \overline{b}$ which implies $\overline{a} \cap \overline{b} = \overline{a}$ (or $\overline{b}$). But if $\overline{a} \cap \overline{b} = \varnothing$ then of course $a \not\sim b$. □

$\overline{a}$ is called the equivalence class of a. By the above proposition we see that we can decompose M into a disjoint union of equivalence classes.

Example C.4. For the integer numbers and the equivalence relation $\mod 2$. We have two equivalence classes:

$$\overline{0} = \{n \mid n \text{ is even}\},$$
$$\overline{1} = \{n \mid n \text{ is odd}\},$$

and $\mathbb{Z} = \overline{0} \cup \overline{1}$ is the disjoint union.

If we choose from every equivalence class $\overline{a}$ a fixed element a_i (this cannot be done in a canonical way!!) then $M = \sqcup_{i \in J} \overline{a_i}$, where J is the index set of disjoint equivalence classes and $\sqcup$ the symbol for disjoint union.

Next we consider the set of equivalence classes

$$M/\sim := \{\overline{a} \mid a \in M\} = \{\overline{a}_i \mid i \in J\}.$$

Note that (1) $\overline{a}$ is considered as an element and not as a subset of M and (2) in the first expression we have to take into account that even if $b \in M$, $a \neq b$ but $\overline{a} = \overline{b}$, the element $\overline{a}$ and $\overline{b}$ will be the same element in the set. Hence they appear only once.
In the previous example, $\mathbb{Z}/\tilde{2} = \{\overline{0}, \overline{1}\} = \{\overline{n} \mid n \in \mathbb{Z}\}$. This set of equivalence classes has exactly two elements.

Remark 13.
1) The presented construction is a very general construction and plays a role everywhere in mathematics and also in computer science. An important technique also in computer science is to identify systems which are "equivalent under certain criteria" and develop for such an "equivalence class of systems" solution. One restricts attention only on the relevant properties ignoring the irrelevant ones. But this is exactly an equivalence relation.

2) Returning back to mathematics: if our set M carries an algebraic structure and if this structure is "compatible" with the equivalence relation, one can endow the set of equivalence classes with the induced structure, called quotient structure. An example was given by $M = \mathbb{Z}$ with $+$ as operation, which defines an abelian group. The equivalence relation $\mod N$ was compatible in the sense that $\overline{a} \oplus \overline{b} := \overline{a + b}$ did not depend on the chosen represented elements $a \in \overline{a}$ and $b \in \overline{b}$. Hence we could also define on $\mathbb{Z}_N$ an induced group structure. This group is called quotient group.

C.2 Coset spaces

Let G be a group (not necessarily abelian) and $H \leqslant G$ a subgroup.
We define $g_1 \underset{H}{\sim} g_2$ to be true if and only if $\exists h \in H$ such that $g_2 = g_1 \cdot h$. For short we use $g_1 \sim g_2$ for $g_1 \underset{H}{\sim} g_2$.

Claim: $\underset{H}{\sim}$ is an equivalence relation for the set G.

Proof. 1. $g \sim g$, take $e \in H$.

2. $g_1 \sim g_2$, then $g_2 = g_1 \cdot h$, $h \in H$. This implies $g_1 = g_2 \cdot h^{-1}$. As H is a subgroup and $h^{-1} \in H$. We get $g_2 \sim g_1$.

3. $g_1 \sim g_2$ and $g_2 \sim g_3$. Hence $\exists h_1, h_2 \in H$ such that $g_2 = g_1 \cdot h_1$ and $g_3 = g_2 \cdot h_2$. We get $g_3 = g_1 \cdot (h_1 \cdot h_2)$. As H is a subgroup, $h_1 \cdot h_2 \in H$. Hence $g_3 \sim g_1$.

This shows $\sim$ is reflexive, symmetric and transitive, hence an equivalence relation. □

This says that we can use the full machinery.

$$\overline{g} := g \mod H = \{g \cdot h \mid h \in H\} = g \cdot H$$

are the equivalence classes and G is the disjoint union of equivalence classes.

In the following, let G be a finite group. In particular

$$G = \bigsqcup_{i=1,\dots,k} g_i H, \qquad g_i \in G$$

is a disjoint union of finitely many inequivalent classes. Let $\{g_i \mid i = 1, \dots, k\}$ be a set of representing elements of the classes $\overline{g}_i$.

Lemma C.5. $\#(g \cdot H) = \#H$.

Proof. We define the map

$$\psi_g : H \mapsto g \cdot H.$$

This map is a bijection.
First we show that it is injective. Let $\psi_g(h_1) = \psi_g(h_2)$. Then $g \cdot h_1 = g \cdot h_2$ and $h_1 = h_2$ as we are in a group.
Next we show surjectivity. Take $k = g \cdot h \in g \cdot H$. Then $k = \psi_g(h)$. □

From

$$G = \bigsqcup_{i=1,\dots,k} g_i H, \qquad g_i \in G$$

we can conclude

$$\#G = \sum_{i=1}^{k} \#(g_i H) = \sum_{i=1}^{k} \#H = (\#H) \cdot k.$$

This shows

Theorem C.6. *(Lagrange)*
Let G be a finite group, $H \leqslant G$ a subgroup. Then $\#H$ divides $\#G$.

Proposition C.7. *Let G be a group with $\#G = p \in \mathbb{P}$ a prime number. Then G does not have any non-trivial subgroups.*

This is evident as $\#H|\#G = p$, hence $\#H = 1$ or $\#H = p$, which says that $H = \{e\}$ or $H = G$.

Proposition C.8. *Let G be a group, $\#G = p$ a prime number. Then G is isomorphic to the cyclic group of order p.*

Proof. As $\#G = p \geq 2$, there exists $a \in G$ such that $a \neq e$. Consider the cyclic subgroup $H =< a >$ generated by a. Then $H \neq \{e\}$. Hence, $G = H$ and G is cyclic. □

From the proof following is immediate:

Proposition C.9. *Given G a group, $\#G = p$ a prime number, then every element $a \in G$, $a \neq e$, is a generator.*

Definition C.10. The gH for $g \in G$ are called cosets (of $G \mod H$). The quotient space $G/_\sim =: G/H$ is called the coset space of $G \mod H$.

Example C.11. Recall our definition of $\mathbb{Z}_2 = \mathbb{Z}/\underset{2}{\sim}$. We can describe it equivalently by considering the group $(\mathbb{Z}, +)$. Then $2\mathbb{Z}$, the even numbers, constitute a subgroup. We consider the general construction of the coset space $\mathbb{Z}/2\mathbb{Z}$ and obtain two classes $\overline{0} = 2\mathbb{Z}$ and $\overline{1} = 1 + 2\mathbb{Z}$. hence $\mathbb{Z}/2\mathbb{Z} = \mathbb{Z}/\underset{2}{\sim}$. In the same way, $\mathbb{Z}/\underset{N}{\sim} = \mathbb{Z}_N$.

As a consequence of the theorem of Lagrange we already concluded earlier:

Proposition C.12. *$ord(a)|\#G$.*

This follows from the fact that $\mathrm{ord}(a) = \# < a >$, the number of elements in the cyclic subgroup generated by a. And this implies

Theorem C.13. *(Little Fermat)*

$$a^{\#G} = e.$$

As $a^{\mathrm{ord}(a)} = e$ and $\#G = k \cdot \mathrm{ord}(a)$, we get $a^{\#G} = a^{\mathrm{ord}(a)\cdot k} = \left(a^{\mathrm{ord}(a)}\right)^k = e^k = e$.

C.3 Quotient groups

Here we like to endow the coset space G/H with a group structure coming from the group structure of G. It will turn out that this will not always be possible. We will need additional properties, either on G or on H.

Case 1

Let $(G, \cdot)$ be a commutative group and H an arbitrary subgroup. We try to make the following definition for the elements of G/H.

$$\overline{g_1}, \overline{g_2} \in G/H : \quad \overline{g_1} \cdot \overline{g_2} := \overline{g_1 \cdot g_2}.$$

In words: we choose representing elements g_1 and g_2 of the classes $\overline{g_1}$ and $\overline{g_2}$ respectively, build the product $g_1 \cdot g_2$ in G and take its class.

Claim: The product is well defined.

Proof. The problem is that by the classes, the elements g_1 and g_2 are not uniquely fixed, e.g. $\overline{g_1} = \overline{g_1'}$, $\overline{g_2} = \overline{g_2'}$ and we have to show that $\overline{g_1 \cdot g_2} = \overline{g_1' \cdot g_2'}$.

$$\overline{g_1'} = \overline{g_1} \quad \leftrightarrow \quad g_1' = g_1 \cdot h_1,$$

$$\overline{g_2'} = \overline{g_2} \quad \leftrightarrow \quad g_2' = g_2 \cdot h_2.$$

We have $g_1' \cdot g_2' = g_1 \cdot h_1 \cdot g_2 \cdot h_2$. As the group G is abelian, we can commute $h_1 \cdot g_2 = g_2 \cdot h_1$ and obtain $g_1' \cdot g_2' = g_1 \cdot g_2 \cdot h_1 \cdot h_2 = g_1 \cdot g_2 \cdot h_3$ with $h_3 \in H$ (as H is a subgroup. Hence $\overline{g_1' \cdot g_2'} = \overline{g_1 \cdot g_2}$. □

Now as everything is defined with the help of the group structure of G, associativity, existence of a neutral element and an inverse is clear.

Theorem C.14. *Let G be a commutative group, H a subgroup. Then G/H is in a natural way also a (commutative) group via $\overline{g_1} \cdot \overline{g_2} := \overline{g_1 \cdot g_2}$. The neutral element is given by $\overline{e}$ (e being the neutral element of G) and $(\overline{g})^{-1} = \overline{g^{-1}}$. G/H is called the quotient group. The natural map*

$$\nu : G \to G/H; g \mapsto \overline{g}$$

is a group homomorphism. Its kernel $\ker \nu$ *is H.*

By its very definition

$$\nu(a \cdot b) = \overline{a \cdot b}$$

and $\nu(a) = \overline{a}$, $\nu(b) = \overline{b}$. But by the definition of the product in G/H, we get $\nu(a \cdot b) = \nu(a) \cdot \nu(b)$, hence it is a group homomorphism, By definition,

$$\ker \nu := \{a \in G \mid v(a) = e\} = \{a \in G \mid \overline{a} = \overline{e}\}.$$

But $\overline{a} = \overline{e} = H \leftrightarrow a \in H$, hence the final claim.

Case 2

We drop the assumption that G is commutative. Above we only used the commutativity at the relation $g_1' \cdot g_2' = g_1 \cdot h_1 \cdot g_2 \cdot h_2$ to commute h_1 with g_2. We could rewrite this as follows

$$g_1' \cdot g_2' = g_1 \cdot g_2 \cdot (g_2^{-1} \cdot h_1 \cdot g_2) \cdot h_2.$$

If we know that $g_2^{-1} \cdot h_1 \cdot g_2 \in H$, $\forall h_1 \in H, \forall g_2 \in G$, then we can conclude as above. If G is abelian, then $g_2^{-1} \cdot h_1 \cdot g_2 = h_1 \in H$. But it is not always true if G is non-abelian.

The following definition already appeared in the main text.

Definition C.15. Let G be an arbitrary group. A subgroup H is called a normal subgroup if for all $g \in G$, we have

$$g^{-1} \cdot H \cdot g = H,$$

i.e. $\forall g \in G$, $\forall h \in H$: $\quad g^{-1} \cdot h \cdot g \in H$.

Hence if H is a normal subgroup, we can argue as above and introduce a well-defined product on G/H:

$$\overline{g_1} \cdot \overline{g_2} = \overline{g_1 \cdot g_2}.$$

Note that if G is abelian, every subgroup is normal. We can generalize the theorem above to:

Theorem C.16. *Let G be a group, H a normal subgroup. Then G/H is in a natural way also a group via $\overline{g_1} \cdot \overline{g_2} := \overline{g_1 \cdot g_2}$. The neutral element is given by $\overline{e}$ (e being the neutral element of G) and $(\overline{g})^{-1} = \overline{g^{-1}}$. G/H is called the quotient group. The natural map*

$$\nu : G \to G/H; g \mapsto \overline{g}$$

is a group homomorphism. Its kernel $\ker \nu$ *is H.*

Remark 14. Let $\varphi : G \to K$ be a group homomorphism. Then

$$\ker \varphi := \{a \in G \mid \varphi(a) = e_K\}$$

is a normal subgroup of G. Imφ is usually only a subgroup of K, not necessarily normal.