Algebraic Structures

Thomas Markwig
Fachbereich Mathematik
Technische Universität Kaiserslautern

Lecture Notes

February 2009

Inhaltsverzeichnis

Introduction

Depending on the degree a participant of the lecture *Algebraic Structures* is aiming at he will take this course in his first or second year. Due to this the audience of the course is rather inhomogeneous. It might either be the first mathematical lecture the student is taking alongside with the lecture *Grundlagen der Mathematik I*, or else he might already be familiar with the mathematical language by one year of practise. I therefore will not start by with a lengthy introduction of the different elements of the mathematical lecture, but I will instead introduce them throughout the lecture whenever they are used first. For a more detailed account I refer the audience to the lecture *Grundlagen der Mathematik I*. I assume that the audience is familiar with the notion of a *set* and basic operations of these like *intersection* or *union*. I also assume that *maps* and simple properties thereof are known, such as *invertibility*. Finally I assume that the *natural numbers* $\mathbb{N}$, the *integers* $\mathbb{Z}$, the *rational numbers* $\mathbb{Q}$, and the *real numbers* $\mathbb{R}$ together with their operations such as *addition* and *multiplication* are familiar.

I should like to notice as well that the lecture and the lecture notes will considerably differ in style. The lecture will be shorter with less text, it will be more graphical in its presentation. Many remarks in the lecture notes will either be completely omitted in the lecture or they will only be given orally. In a sense the lecture and the lecture notes complement each other.

As the title of the course indicates we will study basic algebraic structures such as *groups*, *rings* and *fields* together with maps, which respect the structures. We will spend a lot of time discussing important examples, and I hope to convey thereby their usefulness.

1 Groups and Homomorphisms

In the introduction we said that we wanted to study *structures* on sets. What a *structure* is depends very much on the branch of mathematics we are considering. In this lecture a structure will always consist of one or more *binary operations* on the set, which obey certain *rules*, also called *axioms*. Here a *binary operation* on a set G is a map, which assigns to each pair (g, h) of elements in G again an element in G, i.e. a map $G \times G \longrightarrow G$.

A) Groups

The most basic and at the same time the most important structure on a set is the *group structure*.

Definition 1.1

a. A *group* is a pair $(G, \cdot)$ consisting of a *non-empty* set G and a binary operation "$\cdot$", i.e. a map

$$\cdot : G \times G \to G : (g, h) \mapsto g \cdot h,$$

such that the following *group axioms* are satisfied:

G1: $(g \cdot h) \cdot k = g \cdot (h \cdot k) \quad \forall\, g, h, k \in G,$ ("Associativity")

G2: $\exists\, e \in G \,:\, \forall\, g \in G \,:\, e \cdot g = g,$ ("Existence of a neutral element")

G3: $\forall\, g \in G\ \exists\, g' \in G \,:\, g' \cdot g = e.$ ("Existence of inverse elements")

An element satisfying the property of e is called a *neutral element* of the group G. An element with the property of g' is called an *inverse element of* g.

b. A group is called *abelian* or *commutative* if $(G, \cdot)$ additionally satisfies the following axiom:

G4: $g \cdot h = h \cdot g \quad \forall\, g, h \in G$ ("Commutativity")

c. A group $(G, \cdot)$ is said to be *finite* if $|G| < \infty$, and else it is called *infinite*. $|G|$ is the *order* of G.[1]

Remark 1.2

For several applications the notions of *group* is too restrictive, i.e. we asked for too many axioms. One can weaken the notion in two ways. Suppose for this again that G is a set together with a binary operation "$\cdot$" on G.

a. If the pair $(G, \cdot)$ satisfies only axiom G1 we call $(G, \cdot)$ a *semigroup*.

b. $(G, \cdot)$ is a *monoid*, if only the axioms G1 and G2' are fulfilled:

G2': $\exists\, e \in G \,:\, \forall\, g \in G \,:\, e \cdot g = g \cdot e = g.$ ("Existence of a neutral element")

Note that in the case of a monoid the neutral element has to satisfy a stronger axiom than in the case of a group. Lemma 1.4 will show that in the end the axiom is not really stronger.

[1] $|G|$ denotes the number of elements in the set G.

The notion *abelian*, *finite*, *infinite* and *order* are introduced for semigroups and monoids in the same way as for groups. In this lecture we will not consider properties of semigroups or monoids any further. They were only mentioned for the sake of completeness. □

Before considering further properties of groups it is sensible to give a number of interesting examples to make sure that it is worthwhile spending time on groups.

Example 1.3

a. $(\mathbb{Z},+)$, $(\mathbb{Q},+)$ and $(\mathbb{R},+)$ are abelian groups with the usual addition as operation. In each of the cases the number zero is the neutral element, and for a number g the negative $-g$ is the inverse element.

b. $(\mathbb{Q}\setminus\{0\},\cdot)$ and $(\mathbb{R}\setminus\{0\},\cdot)$ with the usual multiplication as operation are also abelian groups. The number one is in each case the neutral element, and for a number g the inverse is just $\frac{1}{g}$.

c. In contrast to the last example $(\mathbb{Z}\setminus\{0\},\cdot)$ is only an abelian monoid with the number one as neutral element. The axiom G3 is not fulfilled, since only the integers $g=1$ and $g=-1$ have inverses in $\mathbb{Z}\setminus\{0\}$.

d. $(\mathbb{N},+)$ is also only an abelian monoid with the number zero as neutral element, since $g>0$ has no inverse in $\mathbb{N}$.

e. The simplest group is a group containing only one element, $G=\{e\}$, with the group operation defined by $e\cdot e=e$.

Note that in each of the above examples the neutral element was uniquely determined, as was the inverse of an element g. This is no coincidence. The uniqueness follows from the group axioms.

Lemma 1.4
Let $(G,\cdot)$ *be a group.*

a. *The neutral element* $e\in G$ *is uniquely determined and it has the additional property*

$$g\cdot e=g \quad \forall\, g\in G.$$

b. *Let* $g\in G$. *The inverse element* g' *of* g *is uniquely determined and it has the additional property*

$$g\cdot g'=e.$$

The statements formulated in the above lemma can easily be verified in the examples we have considered so far, and we could of course check them for each further example again. This would however be a tedious work. Instead we will deduce the statement in the general form straight from the group axioms. This is then called a *proof* of the statement. Whenever we formulate a statement as a lemma, proposition, theorem or corollary, we will prove them.[2]

Proof of Lemma 1.4: Since we assume that for the pair $(\mathsf{G}, \cdot)$ the axioms G1-G3 are from Definition 1.1 are satisfied, there is a neutral element $e \in \mathsf{G}$, and for an arbitrary but fixed $g \in \mathsf{G}$ there is an inverse element $g' \in \mathsf{G}$.

First we want to show that this e and g' satisfy the additional properties formulated in a. and in b..

Since $(\mathsf{G}, \cdot)$ is a group there must be an element $g'' \in \mathsf{G}$ such that

$$g'' \cdot g' = e. \tag{1}$$

Hence it follows[3]

$$\begin{aligned} g \cdot g' &\overset{G2}{=} e \cdot (g \cdot g') \overset{(1)}{=} (g'' \cdot g') \cdot (g \cdot g') \overset{G1}{=} g'' \cdot (g' \cdot (g \cdot g')) \\ &\overset{G1}{=} g'' \cdot ((g' \cdot g) \cdot g') \overset{G3}{=} g'' \cdot (e \cdot g') \overset{G2}{=} g'' \cdot g' \overset{(1)}{=} e. \end{aligned} \tag{2}$$

This shows that g' satisfies the additional property stated in b., and we therefore get

$$g \cdot e \overset{G3}{=} g \cdot (g' \cdot g) \overset{G1}{=} (g \cdot g') \cdot g \overset{(2)}{=} e \cdot g \overset{G2}{=} g. \tag{3}$$

Since g was an arbitrary element of G, the additional property of e in a. follows as well.

Suppose now that $\tilde{e} \in \mathsf{G}$ is any element with the property of a neutral element of G, i.e.

$$\tilde{e} \cdot h = h \tag{4}$$

for all $h \in \mathsf{G}$. We have to show that $e = \tilde{e}$. Since we already know that e satisfies the additional property in a., we can apply this, i.e. (3), with $\tilde{e}$ in the role of g, and afterwards we can apply (4) with e in the role of h:

$$\tilde{e} \overset{(3)}{=} \tilde{e} \cdot e \overset{(4)}{=} e.$$

Finally we have to show that if $\tilde{g}' \in \mathsf{G}$ is a second inverse element for g, i.e.

$$\tilde{g}' \cdot g = e, \tag{5}$$

[2]The notion *lemma*, *proposition*, *theorem* and *corollary* are usual order structures in mathematics. They contain more or less important statements which we want to remember for further use. In general, the statement of a proposition is considered to be more important than that of a lemma; the statement of a theorem should be more interesting than that of a proposition. A corollary usually follows some proposition or theorem, and its content is then an easy consequence of that proposition or theorem. – In case the proof of some statement would be too advanced for this lecture or we want to skip it for some other reason, then we formulate the statement just as a remark.

[3]Above the equality sign we sometimes give a reason why this equality holds, e.g. from which axiom it follows or a reference to where the equation can be found. This a usual notation in mathematics that we quite frequently use throughout these lecture notes.

then $g' = \tilde{g}'$. For this we apply what we have shown so far:

$$\tilde{g}' \overset{(3)}{=} \tilde{g}' \cdot e \overset{(2)}{=} \tilde{g}' \cdot \left(g \cdot g'\right) \overset{G1}{=} \left(\tilde{g}' \cdot g\right) \cdot g' \overset{(5)}{=} e \cdot g' \overset{G2}{=} g'.$$

Thus all statements of the lemma are proved.[4] $\square$

Notation 1.5

Instead of $(G, \cdot)$ we will often just write G when no ambiguity can occur as far as the operation is concerned. Moreover, we will write gh instead of $g \cdot h$ for $g, h \in G$. The neutral element will sometimes be denoted by 1 instead of e, or by 1_G respectively e_G, if we want to indicate the neutral element of which group it is. The unique inverse element of $g \in G$ will be denoted by g^{-1}, or g_G^{-1} if we want to stress in which group it is the inverse of g.

If the group is abelian, then the operation is often denoted by $+$ instead of $\cdot$. In that case we use the notation 0 respectively 0_G for the neutral element and $-g$ respectively $-g_G$ for the inverse element of $g \in G$. $\square$.

Lemma 1.6

Let $(G, \cdot)$ *be a group,* $g, h, a, b \in G$. *Then the following holds true:*

a. $\left(g^{-1}\right)^{-1} = g$ *and* $(gh)^{-1} = h^{-1}g^{-1}$.

b. *In* G *the* cancellation rules *hold:*
   (i) $ga = gb \ \Rightarrow \ a = b$, *and*
   (ii) $ag = bg \ \Rightarrow \ a = b$.

Proof: a. To prove the first equality, it is enough to show that g satisfies the property of an inverse of g^{-1} in order to show that it is *the* inverse element $(g^{-1})^{-1}$ of g^{-1}, since only one element satisfies that property. For the proof we use the group axioms as well as the additional properties of the inverses shown in Lemma 1.4:

$$g \cdot g^{-1} \overset{\text{Lem. 1.4b.}}{=} e.$$

Hence, g is an inverse of g^{-1}, and by the uniqueness of the inverse element we get as indicated:

$$\left(g^{-1}\right)^{-1} = g.$$

Analogously, by definition $(gh)^{-1}$ is an inverse of gh, and we only have to show that $h^{-1}g^{-1}$ is an inverse of gh as well, in order to show the latter coincides with the former by the uniqueness of inverse elements:

$$\left(h^{-1}g^{-1}\right) \cdot (gh) \overset{G1}{=} h^{-1} \cdot \left(g^{-1} \cdot \left(gh\right)\right) \overset{G1}{=} h^{-1} \cdot \left(\left(g^{-1} \cdot g\right) \cdot h\right)\right)$$
$$\overset{G3}{=} h^{-1} \cdot (e \cdot h) \overset{G2}{=} h^{-1} \cdot h \overset{G3}{=} e.$$

[4] The symbol $\square$ at the *end* of a proof indicates that the proof is finished. Sometimes we use the same symbol at the end of a remark or example to make clear that they are at their end as well.

Hence, $h^{-1}g^{-1}$ is an inverse element of gh, and therefore

$$(gh)^{-1} = h^{-1}g^{-1}.$$

b. The first cancellation rule follows by multiplication with the inverse of g from the left:[5]

$$a \overset{G2}{=} e \cdot a \overset{G3}{=} \left(g^{-1} \cdot g\right) \cdot a \overset{G1}{=} g^{-1} \cdot (g \cdot a)$$
$$\overset{Ass.}{=} g^{-1} \cdot (g \cdot b) \overset{G1}{=} \left(g^{-1} \cdot g\right) \cdot b \overset{G3}{=} e \cdot b \overset{G2}{=} b.$$

Similarly the second cancellation rule holds by multiplication with g^{-1} from the right and by the additional property of inverses from Lemma 1.4. The details are left to the reader.

□

Remark 1.7
So far whenever we applied the associativity rule several times we did so step by step. From now on, however, we will rearrange brackets freely, i.e. combine several applications of the associativity rule in one. This should not lead to any confusion. □

We are used to write repeated multiplications of a number with itself with the aid of powers, and we will introduce the analogous notion now for arbitrary groups.

Definition 1.8
Let $(G, \cdot)$ be a group, $g \in G$. We set $g^0 := e$, and for $n \in \mathbb{Z}$, $n > 0$, we define recursively $g^n := g \cdot g^{n-1}$, and finally we set $g^{-n} := (g^{-1})^n$.

For this definition we used a property of the natural numbers which is well-known to everyone, and which we call the *principle of induction*:

> *Starting from a natural number n_0 we can get every larger natural number by repeatedly adding the number 1.*

In an axiomatic treatment of the natural numbers this is one of their axioms. For us, however, it is just a familiar property. We actually have already used this property in the above *recursive* definition of g^n for non-negative natural numbers n. We first defined it for $n = 0$, and we then reduced the definition of g^n to the definition of g^{n-1}.

One very often uses this property of the natural numbers as a technique to prove a statement $\mathcal{A}$

- which depend on a natural number n, and
- which we want to prove for all $n \geq n_0$.

[5] The abbreviation *Ass.* above the equality sign indicates that the equality holds *by assumption*. In this particular case we assumed that $ga = gb$ holds.

The dependence of the statement $\mathcal{A}$ on the natural number n is expressed by appending n as an index, i.e. we write $\mathcal{A}_n$ instead of $\mathcal{A}$. A typical example for such a statement would be:

$$\mathcal{A}_n : \textit{a number of the form } n^3 - n \textit{ is divisible by } 6$$

where $n \in \mathbb{N}$ is any natural number, i.e. $n_0 = 0$ in this example. If we want to prove that this statement $\mathcal{A}_n$ holds for every $n \geq n_0$, we show first that it holds for n_0 (we call this the *induction basis*); afterwards we show that if $\mathcal{A}_n$ holds for some arbitrary but fixed number n (to assume that it holds for some n is called the *induction hypothesis*), then it also holds for $n + 1$ (this is called the *induction step*). The above described property of the natural numbers allows then to deduce from $\mathcal{A}_{n_0}$ the correctness of $\mathcal{A}_{n_0+1}$, and then the correctness of $\mathcal{A}_{n_0+2}$ and going on like this the correctness of $\mathcal{A}_n$ for every natural number $n \geq n_0$. We leave it to the reader to prove the statement in our example by the principle of induction. Instead we formulate the principle in a more compact way.

Remark 1.9 (Principle of Induction)
Let a statement $\mathcal{A}_n$ be correct for some natural number $n = n_0$ (*induction basis*), and in addition the following shall hold true: if the statement holds for some $n \geq n_0$ (*induction assumption*), is also holds for $n+1$ (*induction step*). Then the statement holds for all natural numbers $n \geq n_0$. $\square$

If we use this technique to prove statements which are indexed by n, then we say that we do *induction on* n. We will now use this technique to prove the *power laws.*

Lemma 1.10 (Power Laws)
Let $(G, \cdot)$ *be a group,* $g \in G$*,* $n, m \in \mathbb{Z}$*, then the following power laws hold true:*

$$g^n \cdot g^m = g^{n+m} \qquad \textit{and} \qquad (g^m)^n = g^{m \cdot n}.$$

Proof: We first want to show that

$$g^n = (g^{-1})^{-n}. \tag{6}$$

If $n < 0$ this holds by definition. If $n > 0$, then $-n < 0$ and by definition and Lemma 1.6 we have[6]

$$(g^{-1})^{-n} \overset{\text{Def.}}{=} \left((g^{-1})^{-1}\right)^{-(-n)} \overset{\text{Lem. 1.6}}{=} g^n.$$

And finally, if $n = 0$ then $g^n = e = (g^{-1})^{-n}$ by definition. This shows (6).

We next want to consider a special case of the first power law, namely

$$g \cdot g^n = g^{n+1} \tag{7}$$

for all natural numbers $n \in \mathbb{N}$. If $n \geq 0$ this follows from the definition; if $n < 0$ then $-n - 1 \geq 0$ and by definition we get

$$\left(g^{-1}\right) \cdot g^{n+1} \overset{(6)}{=} \left(g^{-1}\right) \cdot \left(g^{-1}\right)^{-n-1} \overset{\text{Def}}{=} \left(g^{-1}\right)^{-n-1+1} = \left(g^{-1}\right)^{-n} \overset{(6)}{=} g^n. \tag{8}$$

[6]The abbreviation *Def.* above the equality sign means that the equality holds *by definition.*

If we multiply both sides of the equation by g, we get

$$g \cdot g^n \overset{(8)}{=} g \cdot \left(g^{-1} \cdot g^{n+1}\right) \overset{G1}{=} \left(g \cdot g^{-1}\right) \cdot g^{n+1} \overset{G3}{=} e_G \cdot g^{n+1} \overset{G2}{=} g^{n+1}.$$

Thus we have proved (7).

We now come to the general rule

$$g^n \cdot g^m = g^{n+m} \tag{9}$$

for $n, m \in \mathbb{Z}$, and we do the proof by induction considering different cases.

<u>1st Case: $n \geq 0$.</u> We want to prove this case by *induction on* n. Formulated more precisely, we have to show for an arbitrary but fixed $g \in G$ and $m \in \mathbb{Z}$ that the statement

$$\mathcal{A}_n: \qquad g^n \cdot g^m = g^{n+m}$$

holds for all $n \in \mathbb{N}$. For this we have to check first the *induction basis*,i.e. that the statement holds for some starting number – in Remark 1.9 this was the number n_0, here it is the number 0.

If $n = 0$ then

$$g^n \cdot g^m \overset{n=0}{=} g^0 \cdot g^m \overset{\text{Def.}}{=} e \cdot g^m \overset{G2}{=} g^m \overset{n=0}{=} g^{n+m}.$$

This proves the induction basis $\mathcal{A}_0$. We next have to do the *induction step*, i.e. whenever the statement holds for a number n it also holds for the next larger number $n+1$. For this we may assume that the *induction hypothesis* holds, i.e. that $\mathcal{A}_n$ for the given $n \geq n_0$ is correct, and we have to deduce that then also $\mathcal{A}_{n+1}$ is correct. By definition and induction hypothesis we have[7]

$$g^{n+1} \cdot g^m \overset{\text{Def.}}{=} (g \cdot g^n) \cdot g^m \overset{G1}{=} g \cdot (g^n \cdot g^m) \overset{\text{Ind.}}{=} g \cdot g^{n+m} \overset{(7)}{=} g^{n+1+m}.$$

Thus we have deduced from the correctness of $\mathcal{A}_n$ the correctness of $\mathcal{A}_{n+1}$. The principle of induction therefore allows us to deduce that the statement holds for all n which are at least 0. It remains to consider the case $n < 0$.

<u>2nd Case: $n < 0$.</u> The first case (applied to g^{-1} and $-m$) implies by the fact that $-n > 0$:

$$g^n \cdot g^m \overset{(6)}{=} (g^{-1})^{-n} \cdot (g^{-1})^{-m} \overset{1\text{stCase}}{=} (g^{-1})^{-n-m} \overset{(6)}{=} g^{n+m}.$$

This finishes the proof of the first power law (9). We now want deduce from this the special case

$$(g^n)^{-1} = g^{-n}. \tag{10}$$

of the second power law. $g^{-n} \cdot g^n = g^{-n+n} = g^0 = e$ implies that g^{-n} is an inverse of g^n. The correctness of (10) thus follows from the uniqueness of the inverse.

We are now able to prove the second power law,

$$(g^m)^n = g^{m \cdot n}$$

[7] The abbreviation *Ind.* over an equality sign means, that this equality holds by the *induction hypothesis*, i.e. since the statement $\mathcal{A}_n$ holds true. In our concrete example this means, that for fixed m, n and g the equality $g^n \cdot g^m = g^{n+m}$ holds.

for $m, n \in \mathbb{Z}$. We do so by considering again different cases.[8]

<u>1st Case: $n \geq 0$.</u> We want to show for arbitrary but fixed $g \in G$ and $m \in \mathbb{Z}$ by induction on n the statement

$$\mathcal{A}_n: \qquad (g^m)^n = g^{m \cdot n}.$$

<u>$n = 0$:</u> Then

$$(g^m)^n \overset{n=0}{=} (g^m)^0 \overset{\text{Def.}}{=} e \overset{\text{Def.}}{=} g^0 \overset{\text{Def.}}{=} g^{m \cdot n}.$$

<u>$n \mapsto n+1$:</u> By definition, induction hypothesis and the 2nd power law we have:

$$(g^m)^{n+1} \overset{\text{Def.}}{=} (g^m) \cdot (g^m)^n \overset{\text{Ind.}}{=} g^m \cdot g^{m \cdot n} \overset{(9)}{=} g^{m+m \cdot n} \overset{\text{Def.}}{=} g^{m \cdot (n+1)}.$$

<u>2nd Case: $n < 0$.</u> The 1st Case implies by $-n > 0$:

$$(g^m)^n \overset{(6)}{=} \left((g^m)^{-1}\right)^{-n} \overset{(10)}{=} (g^{-m})^{-n} \overset{\text{1stCase}}{=} g^{(-m)\cdot(-n)} \overset{\text{Def.}}{=} g^{m \cdot n}.$$

□

Remark 1.11

Is $(H, \cdot)$ a semigroup (or monoid) and $g \in H$ we define for $0 \neq n \in \mathbb{N}$ (resp. $n \in \mathbb{N}$) the element g^n analogously and show that for $0 \neq n, m \in \mathbb{N}$ (resp. $n, m \in \mathbb{N}$) the above power laws hold as well. The proofs are the same. □

Remark 1.12

So far we explained every step in our proves in detail and we justified each and every equality by some reference. By now the reader should have understood the principle, and we will be less detailed for the remaining part of the lecture notes. □

All examples considered in Example 1.3 are abelian, so that the inversion rule "$(gh)^{-1} = h^{-1}g^{-1}$" takes the form "$(gh)^{-1} = g^{-1}h^{-1}$". In order to see that this does not hold in general, we need an example of a non-abelian group. For this we recall the notion of a *bijective map* and the composition of such maps.

Definition 1.13

For a non-empty set M we define

$$\text{Sym}(M) := \{f : M \to M \mid f \text{ is bijective}\}.$$

The composition $\circ$ defines a binary operation on $\text{Sym}(M)$, and we call the pair $(\text{Sym}(M), \circ)$ the *symmetric group on* M. The elements of $\text{Sym}(M)$ well be called *permutations* of M.

---

[8] The first case $n \geq 0$ will yet again be proved by induction on n. However we will considerably shorten our notation. We know by now that we have to check the statement in question for some initial value n_0 (*induction basis*) and that we then have to deduce the correctness of the statement for $n+1$ (*induction step*) under the assumption that it is correct for n (*induction assumption*). We will indicate this as <u>$n = 0$</u> for the *induction basis* and as <u>$n \mapsto n+1$</u> for the *induction step*. It might take a while to get used to this short hand notation, but in the end it has the big advantage that due to its shortness it does not distract from the essentials by too many words.

For $M = \{1, \ldots, n\}$ we write $\mathbb{S}_n$ instead of $\mathrm{Sym}(M)$ and we call $\mathbb{S}_n$ the *symmetric group of degree* n.

The notion symmetric *group* is only justified, if we show that $(\mathrm{Sym}(M), \circ)$ is indeed a group.

Proposition 1.14
$(\mathrm{Sym}(M), \circ)$ *is a group, and it is abelian if and only if* $|M| \leq 2$. *The neutral elements is* id_M*, and the inverse element of a map* $f \in \mathrm{Sym}(M)$ *is its inverse mapping.*

Proof: We first want to show that the composition of two bijective maps is bijective again, i.e. that "$\circ$" is indeed a binary operation on $\mathrm{Sym}(M)$.

Let $f, g : M \to M$ be bijective. Then there exist inverse mappings $f^{-1} : M \to M$ and $g^{-1} : M \to M$, and using the associativity of the composition we get

$$(f \circ g) \circ \left(g^{-1} \circ f^{-1}\right) = f \circ \left(g \circ g^{-1}\right) \circ f^{-1} = f \circ \mathrm{id}_M \circ f^{-1} = f \circ f^{-1} = \mathrm{id}_M,$$

and analogously $\left(g^{-1} \circ f^{-1}\right) \circ (f \circ g) = \mathrm{id}_M$. Thus $f \circ g$ is bijective.

The associativity of "$\circ$", i.e. the axiom G1, follows since the composition of maps is known to be associative. The identity map id_M on M is bijective and has the property that $\mathrm{id}_M \circ f = f$ for all $f \in \mathrm{Sym}(M)$. Thus it is the neutral element of $(\mathrm{Sym}(M), \circ)$. The inverse mapping of $f \in \mathrm{Sym}(M)$ is an inverse element of f in the sense of axiom G3. Hence, $(\mathrm{Sym}(M), \circ)$ is a group.

It remains to show:

$$(\mathrm{Sym}(M), \circ) \text{ is abelian} \quad \Longleftrightarrow \quad |M| \leq 2.$$

If $|M| \geq 3$ we can choose three pairwise different elements[9] $m, m', m'' \in M$ and we can define the maps

$$f : M \longrightarrow M : n \mapsto \begin{cases} m', & \text{if } n = m, \\ m, & \text{if } n = m', \\ n, & \text{else,} \end{cases}$$

and

$$g : M \longrightarrow M : n \mapsto \begin{cases} m'', & \text{if } n = m, \\ m, & \text{if } n = m'', \\ n, & \text{else.} \end{cases}$$

Obviously, $f \circ f = \mathrm{id}_M$ and $g \circ g = \mathrm{id}_M$, so that f and g are bijective with inverses $f^{-1} = f$ and $g^{-1} = g$ respectively. Moreover by definition

$$(f \circ g)(m) = f\big(g(m)\big) = f(m'') = m'',$$

but

$$(g \circ f)(m) = g\big(f(m)\big) = g(m') = m' \neq m''.$$

[9] We call the three elements *pairwise different* if $m \neq m'$, $m' \neq m''$ and $m'' \neq m$, i.e. if each pair of the the three elements consists of different elements. This notion generalises in an obvious way to sets with an arbitrary number of elements.

This shows that $g \circ f \neq f \circ g$, and $(\mathrm{Sym}(M), \circ)$ is not abelian. This proves the implication "$\Longrightarrow$" in our statement by *contraposition*.[10] The other implication "$\Longleftarrow$" is left to the reader as an exercise. $\square$

Exercise 1.15
Check if the following binary operation on the set $G := \mathbb{Q} \times \mathbb{Q}$ makes G a group:

$$G \times G \longrightarrow G : \big((a,b),(a',b')\big) \mapsto (a,b) \cdot (a',b') := (aa', bb').$$

Exercise 1.16
Check if the following binary operation on the set $G := \mathbb{Q}_{>0} \times \mathbb{Q}_{>0}$ makes G a group:

$$G \times G \longrightarrow G : \big((a,b),(a',b')\big) \mapsto (a,b) \cdot (a',b') := (aa', bb').$$

Exercise 1.17
Let $(G, \cdot)$ and $(H, *)$ be two groups. We define on the set $G \times H = \{(x,y) \mid x \in G, y \in H\}$ a binary operation by

$$(x,y) \circ (x',y') := (x \cdot x', y * y')$$

for $(x,y), (x',y') \in G \times H$. Show that $(G \times H, \circ)$ is a group.

Exercise 1.18
Check which of the following binary operations defines a group structure:

a. $G = \big(\mathbb{Q}\setminus\{0\}\big) \times \big(\mathbb{Q}\setminus\{0\}\big)$ mit $(a,b)\cdot(a',b') = (ab', ba')$ for $a, a', b, b' \in \mathbb{Q}\setminus\{0\}$,
b. $G = \mathbb{R}\times\mathbb{R}\setminus\big\{(0,0)\big\}$ mit $(a,b)\cdot(a',b') = (aa'-bb', ab'+ba')$ for $a, a', b, b' \in \mathbb{R}$.

Exercise 1.19
Find all possible binary operations on the set $G = \{e, a, b\}$ such that G is a group with neutral element e.

Exercise 1.20
Find all possible binary operations on the set $G = \{e, a, b, c\}$ for which G is a group with neutral element e. List only those examples which cannot be transformed into each other by just permuting the letters a, b and c.

Exercise 1.21
A scheme of the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $a, b, c, d \in \mathbb{R}$ is a *real* 2x2-*matrix*, and $\mathrm{Mat}_2(\mathbb{R})$ is the set of all such matrices. For two real 2x2-matrices we define their product as

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

[10] To prove an implication "$A \Longrightarrow B$" by *contraposition* means that one instead shows the implication "$\neg B \Longrightarrow \neg A$". Both implications are equivalent to each other, i.e. one holds true if and only if the other one does.

Moreover we define

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \in \mathbb{R}$$

and call it the *determinant* of the matrix. Finally, we set

$$\mathrm{Gl}_2(\mathbb{R}) = \{A \in \mathrm{Mat}_2(\mathbb{R}) \mid \det(A) \neq 0\}.$$

Show:

a. For $A, B \in \mathrm{Mat}_2(\mathbb{R})$ gilt $\det(A \cdot B) = \det(A) \cdot \det(B)$.

b. $\big(\mathrm{Gl}_2(\mathbb{R}), \cdot\big)$ is a non-abelian group.

Exercise 1.22
Let $(G, \cdot)$ be a group and $a \in G$ be fixed. We define a binary operation on G by

$$* : G \times G \longrightarrow G : (g, h) \mapsto g * h = g \cdot \big(a^{-1} \cdot h\big).$$

Check if $(G, *)$ is a group.

Exercise 1.23 (Boolean Group)
Let M be a set.

a. If $X, Y, Z \subseteq M$, then

$$X \setminus \big((Y \setminus Z) \cup (Z \setminus Y)\big) = \big(X \setminus (Y \cup Z)\big) \cup (X \cap Y \cap Z)$$

and

$$\big((X \setminus Y) \cup (Y \setminus X)\big) \setminus Z = \big(X \setminus (Y \cup Z)\big) \cup \big(Y \setminus (X \cup Z)\big).$$

b. We define on the power set $G = \mathcal{P}(M) = \{A \mid A \subseteq M\}$ of M a binary operation Operation by

$$A + B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$$

for $A, B \in G$. Show that $(G, +)$ is an abelian group.

Exercise 1.24
Let $(G, \cdot)$ be a group with neutral element e. If $g^2 = e$ for all $g \in G$ then G is abelian.

Exercise 1.25
Let M be a set, $m \in M$, $k \in \mathbb{N}$ and $\sigma \in \mathrm{Sym}(M)$ with $\sigma^k(m) = m$. Then $\sigma^{q \cdot k}(m) = m$ for all $q \in \mathbb{Z}$.

B) **Subgroups**

An important principle in mathematics is to consider for each structure so called *substructures*. For a set these are the subsets, for a group these will be the subgroups, i.e. subsets which *respect* the group structure. A group consists of a set G together with a binary operation $\cdot : G \times G \to G$ which satisfies certain axioms. If $U \subseteq G$ is a subset, then we can restrict the map "$\cdot$" to $U \times U$. That way we get a map

$$U \times U \longrightarrow G : (u, v) \mapsto u \cdot v,$$

where $u \cdot v$ is some element in G which in general will not be an element of U again. The latter means that the restriction of "$\cdot$" to $U \times U$ will in general not be a binary operation on U! However, this is certainly a minimal requirement if we want to say that U respects the group structure "$\cdot$". Let us now suppose that against all hope the restriction of "$\cdot$" actually defines a binary operation on U. It then is natural to ask, if U with this binary operation satisfies the axioms G1-G3, i.e. if it is a group. If so, we can really say that U respects the group structure on G. These considerations lead to the notion of a *subgroup*, and similar considerations work for all algebraic structures.

Definition 1.26
Let $(G, \cdot)$ be a group. A subset $U \subseteq G$ is a *subgroup* of G, if

$$u \cdot v \in U \quad \text{for all } u, v \in U$$

and if moreover $(U, \cdot)$ is a group, i. e. if the restriction of the operation "$\cdot$" to $U \times U$ makes U a group.

Notation 1.27
Is $(G, \cdot)$ a group and $U \subseteq G$, we will use the notation $U \leq G$ to indicate that U is a subgroup of $(G, \cdot)$.

Before we consider examples of subgroups we want to formulate and prove a criterion which makes it simpler to check if a certain subset is actually a subgroup.

Proposition 1.28 (Criterion for Subgroups)
Let $(G, \cdot)$ be a group and $\emptyset \neq U \subseteq G$ a non-empty subset. The following statements are equivalent:

a. U *is a subgroup of* G*,*

b. $\forall\, u, v \in U : uv \in U$ *and* $u_G^{-1} \in U$*.*

The properties in b. are known as the closedness *of* U *with respect to the group operation and with respect to inverses.*

Proof: "a. $\Rightarrow$ b.": Suppose first that U is a subgroup of G. The image of $U \times U$ by the map "$\cdot$" is then contained in U by definition, i.e. for all $u, v \in U$ we have $uv \in U$. Moreover, U satisfies the group axioms. Let $e_U \in U$ be the neutral element of U and $e_G \in G$ be the neutral element of G. Moreover, for an element $u \in U \subseteq G$ we denote by u_G^{-1} its inverse in G and by u_U^{-1} its inverse in U, i.e. $u_G^{-1}u = uu_G^{-1} = e_G$ and $u_U^{-1}u = uu_U^{-1} = e_U$. In the following equation we need the inverse of the element e_U in the group G, which leads to the clumsy notation $(e_U)_G^{-1}$. With these conventions we get:

$$e_U \stackrel{\text{G2 in G}}{=} e_G e_U \stackrel{\text{G3 in G}}{=} \big((e_U)_G^{-1} e_U\big) e_U \stackrel{\text{G1 in G}}{=} (e_U)_G^{-1}(e_U e_U) \stackrel{\text{G2 in U}}{=} (e_U)_G^{-1} e_U \stackrel{\text{G3 in G}}{=} e_G. \tag{11}$$

Moreover, we have

$$u_U^{-1}u \stackrel{\text{G3 in U}}{=} e_U \stackrel{(11)}{=} e_G.$$

Thus $u_U^{-1} = u_G^{-1}$ by the uniqueness of the inverse element in G. This in particular implies that $u_G^{-1} \in U$.

<u>"a. ⇐ b.":</u> Since $uv \in U$ for all $u, v \in U$, the image of $U \times U$ by the map "$\cdot$" is indeed contained in U. It remains to show that the axioms G1-G3 are satisfied. Note first that the axiom G1 on U is inherited from G since all the elements of U are contained in G. Since U is non-empty there exists an element $u \in U$. By assumption its inverse element $u_G^{-1} \in U$ is contained in U and therefore $e_G = u_G^{-1}u \in U$. Since $e_G u = u$ for all $u \in U$, the axiom G2 is satisfied and we have $e_U = e_G$. Moreover, since for $u \in U$ also $u_G^{-1} \in U$ and since

$$u_G^{-1} \cdot u = e_G = e_U,$$

also G3 is satisfied, and the inverse of u in U coincides with the inverse of u in G. □

The statement of the following corollary was part of the proof of the criterion for subgroups.

Corollary 1.29

Is $(G, \cdot)$ a group and $U \leq G$, then the neutral element e_U of the group $(U, \cdot)$ and the neutral element e_G of the group $(G, \cdot)$ coincide. Moreover, for each $u \in U$ the inverse u_U^{-1} of u in $(U, \cdot)$ coincides with the inverse u_G^{-1} of u in $(G, \cdot)$.

We now want to apply the criterion for subgroups in order to find subgroups of the previously considered groups.

Example 1.30

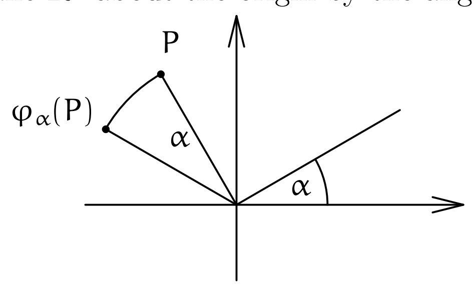
a. Is $(G, \cdot)$ a group with neutral element e_G, then the subsets $\{e_G\}$ and G of G are always groups. They are called the *trivial subgroups.*

b. $(\{-1, 1\}, \cdot)$ is a subgroup subgroup of $(\mathbb{Q} \setminus \{0\}, \cdot)$, as follows immediately from Proposition 1.28.

c. For $\alpha \in \mathbb{R}$ the map

$$\varphi_\alpha : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \to \big(\cos(\alpha) \cdot x - \sin(\alpha) \cdot y, \sin(\alpha) \cdot x + \cos(\alpha) \cdot y\big)$$

is a rotation in the plane $\mathbb{R}^2$ about the origin by the angle α.

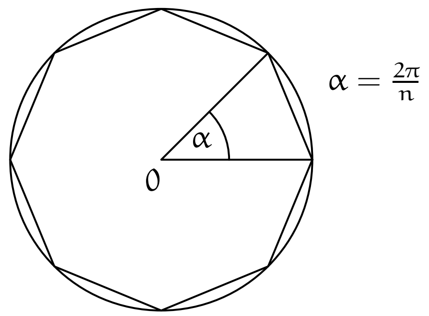


Obviously we have $\varphi_\alpha \circ \varphi_\beta = \varphi_{\alpha+\beta}$ for $\alpha, \beta \in \mathbb{R}$, and for $\alpha \in \mathbb{R}$ the inverse is thus $\varphi_{-\alpha} = (\varphi_\alpha)^{-1}$, since $\varphi_0 = \mathrm{id}_{\mathbb{R}^2}$. In particular φ_α is bijective for each $\alpha \in \mathbb{R}$. Proposition 1.28 thus implies that the set

$$\mathrm{SO}(2) := \{\varphi_\alpha : \mathbb{R}^2 \to \mathbb{R}^2 \mid \alpha \in \mathbb{R}\}$$

is a subgroup of $\mathrm{Sym}\left(\mathbb{R}^2\right)$.

d. Let $E_n \subset \mathbb{R}^2$ be the *regular* n*-gon*.



We set

$$U := \left\{\varphi_\alpha \in \mathrm{SO}(2) \;\middle|\; \varphi_\alpha(E_n) = E_n\right\}.$$

Claim: $(U, \circ)$ is a subgroup of $\big(\mathrm{SO}(2), \circ\big)$.

For $\varphi_\alpha, \varphi_\beta \in U$ we have

$$(\varphi_\alpha \circ \varphi_\beta)(E_n) = \varphi_\alpha\big(\varphi_\beta(E_n)\big) = \varphi_\alpha(E_n) = E_n$$

and

$$\varphi_\alpha^{-1}(E_n) = \varphi_\alpha^{-1}\big(\varphi_\alpha(E_n)\big) = \big(\varphi_\alpha^{-1} \circ \varphi_\alpha\big)(E_n) = \mathrm{id}_{\mathbb{R}^2}(E_n) = E_n.$$

Moreover, $\varphi_\alpha \circ \varphi_\beta \in U$ and $\varphi_\alpha^{-1} \in U$. And since $\mathrm{id}_{\mathbb{R}^2} = \varphi_0 \in U$, we have $U \neq \emptyset$, so that U is a subgroup of $\mathrm{SO}(2)$ by Proposition 1.28.

One checks easily that U consists of the rotations φ_α with $\alpha = k \cdot \frac{2\pi}{n}$, $k = 0, \ldots, n-1$. In particular, $|U| = n$.

e. Let $n \in \mathbb{Z}$ and $n\mathbb{Z} := \{nz \mid z \in \mathbb{Z}\}$ be the set of all multiples of n

Claim: $(n\mathbb{Z}, +)$ is a subgroup of $(\mathbb{Z}, +)$.

If $nz, nz' \in n\mathbb{Z}$ then $nz + nz' = n(z + z') \in n\mathbb{Z}$ and $-(nz) = n \cdot (-z) \in n\mathbb{Z}$. Moreover, $\emptyset \neq n\mathbb{Z} \subset \mathbb{Z}$, since $0 = n \cdot 0 \in n\mathbb{Z}$. Thus again Proposition 1.28 implies the claim.

f. For two integers $m, n \in \mathbb{Z}$ we have $m\mathbb{Z} \subseteq n\mathbb{Z}$ if and only if m is a multiple n.

g. The inclusion $\mathbb{Z} \subset \mathbb{Q}$, $\mathbb{Z} \subset \mathbb{R}$ and $\mathbb{Q} \subset \mathbb{R}$ makes the subsets subgroups with respect to the addition as group structure.

Remark 1.31

How do subgroups behave with respect to set operations like the union?

Let's have a look at the group $(\mathbb{Z}, +)$ and its subgroups $2\mathbb{Z} \leq \mathbb{Z}$ and $3\mathbb{Z} \leq \mathbb{Z}$. The set $U = 2\mathbb{Z} \cup 3\mathbb{Z}$ is not closed with respect to the group operation, since

$$2 + 3 = 5 \notin 2\mathbb{Z} \cup 3\mathbb{Z},$$

since 5 is neither divisible by 2 nor by 3. Thus the union of subgroups is not a subgroup in general.

In contrast to the union the intersection of subgroups is always a subgroup.

Lemma 1.32
Let $(G, \cdot)$ *be a group,* I *some index set and* $U_i \leq G$ *for* $i \in I$*. Then*

$$\bigcap_{i \in I} U_i \leq G.$$

Proof: The proof is left to the reader as an easy application of the criterion for subgroups. □

We now use this property to repair the flaw that subgroups do not behave well with respect to unions. For this we replace the notion of the union of two subgroups by the smallest subgroup which contains both. For this we need the subgroup *generated* by some set.

Definition 1.33
Let $(G, \cdot)$ be a group and $M \subseteq G$ any subset. The subgroup *generated* by M defined as

$$\langle M \rangle = \bigcap_{M \subseteq U \leq G} U,$$

i.e. it is the intersection of all subgroups of G, which contain M. For $M = \{g_1, \ldots, g_n\}$ we write in general only $\langle g_1, \ldots, g_n \rangle$ instead of $\langle \{g_1, \ldots, g_n\} \rangle$.

Such a definition is useful, since we get the fact that it is a subgroup for free and also that it is the smallest subgroup which contains M. However, the definition gives no hint how the elements in $\langle M \rangle$ look like. Fortunately, this is completely described by the following proposition, which also explains in which sense the elements of M actually *generate* the subgroup.

Proposition 1.34
Let $(G, \cdot)$ *be a group and* $M \subseteq G$ *be any subset. Then*[11]

$$\langle M \rangle = \left\{ g_1^{\alpha_1} \cdots g_n^{\alpha_n} \mid n \geq 0, g_1, \ldots, g_n \in M, \alpha_1, \ldots, \alpha_n \in \mathbb{Z} \right\}.$$

Proof: Let us introduce a name for the right hand side of the equation:

$$N = \{ g_1^{\alpha_1} \cdots g_n^{\alpha_n} \mid n \geq 0, g_1, \ldots, g_n \in M, \alpha_1, \ldots, \alpha_n \in \mathbb{Z} \big\},$$

We first want to show that $N \subseteq \langle M \rangle$. If $U \leq G$ such that $M \subseteq U$, then $g_1^{\alpha_1} \cdots g_n^{\alpha_n} \in U$ for all $g_i \in M$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}$. Thus $N \subseteq U$, and hence $N \subseteq \langle M \rangle$.

It remains to show that $\langle M \rangle \subseteq N$. For this it suffices to show that $N \leq G$ with $M \subseteq N$. Since the empty product by convention is the neutral element e_G the set N is non-empty. Let now $h = g_1^{\alpha_1} \cdots g_n^{\alpha_n}, h' = g_{n+1}^{\alpha_{n+1}} \cdots g_m^{\alpha_m} \in N$ be two arbitrary elements in N, then

$$h \cdot h' = g_1^{\alpha_1} \cdots g_m^{\alpha_m} \in N$$

and

$$h^{-1} = g_n^{-\alpha_n} \cdots g_1^{-\alpha_1} \in N.$$

[11]Note that for $n = 0$ we have the empty product which by definition is just e_G.

Hence $N \leq G$ is a subgroup of G, and since $M \subseteq N$ the claim follows. $\square$

Example 1.35
Is $(G, \cdot)$ a group and $g \in G$, then Proposition 1.34 implies

$$\langle g \rangle = \{g^k \mid k \in \mathbb{Z}\}.$$

If we apply this to the group $(\mathbb{Z}, +)$ and a number $n \in \mathbb{Z}$, then the subgroup generated by $M = \{n\}$ is the subgroup

$$n\mathbb{Z} = \{n \cdot z \mid z \in \mathbb{Z}\} = \langle n \rangle = \langle M \rangle.$$

Definition 1.36
A group $(G, \cdot)$ is called *cyclic*, if it is generated by a single element, i.e. if there is a $g \in G$ such that $G = \langle g \rangle$.

For the proof of the next proposition we need the principle of *division with remainder* in the integers, a property of the integers which we suppose to be well known. One could prove it by an induction with an extensive case distinction. We, moreover, use the *Archimedian principle* of the integers whose content is equally well known. For the sake of completeness we formulate both principles here.

Remark 1.37 (Division with Remainder)
For integers $m, n \in \mathbb{Z}$ with $n \neq 0$ there exist uniquely determined integers $q, r \in \mathbb{Z}$ such that

$$m = qn + r \quad \text{and} \quad 0 \leq r < |n|. \tag{12}$$

We call r the *remainder of* m *modulo* n.

Remark 1.38 (Archimedian Principle)
Every non-empty set of natural numbers contains a smallest element.

Proposition 1.39
$U \subseteq \mathbb{Z}$ *is a subgroup of* $(\mathbb{Z}, +)$ *if and only if there is an integer* $n \geq 0$ *such that* $U = n\mathbb{Z} = \langle n \rangle$*. In particular every subgroup of* $(\mathbb{Z}, +)$ *is cyclic.*

Proof: From Example 1.30 and Example 1.35 we know that the sets of the form $n\mathbb{Z} = \langle n \rangle$ are subgroups of $(\mathbb{Z}, +)$.

Let thus $U \leq \mathbb{Z}$ be a subgroup of $(\mathbb{Z}, +)$. It remains to show that there is an integer $n \geq 0$ such that $U = n\mathbb{Z}$. If $U = \{0\}$ we can choose $n = 0$. If $U \neq \{0\}$ there is an integer $0 \neq z \in U$ and one of the integers $z \in U$ or $-z \in U$ is positive. Hence the subset

$$\{m \in \mathbb{N} \mid 0 \neq m \in U\}$$

of the natural numbers is non-empty and contains therefore by the Archimedian principle a smallest element, say

$$n := \min\{z \in U \mid z > 0\} \in U.$$

We want to show that $U = \langle n \rangle$. Since $n \in U$ we get right away that $\langle n \rangle \subseteq U$ by the definition of $\langle n \rangle$. Let vice versa $u \in U$. Division with remainder of u by n gives integers $q, r \in \mathbb{Z}$ such that

$$u = q \cdot n + r$$

and

$$0 \leq r < n. \tag{13}$$

Since u and n are both elements of U, also

$$r = u - q \cdot n \in U.$$

But since n is the smallest positive integer in U and since $r < n$ is non-negative, we conclude that r must be zero by (13). Thus $u = q \cdot n \in \langle n \rangle$, and we have proved $U = \langle n \rangle$. □

Exercise 1.40
Prove Lemma 1.32.

Exercise 1.41
Show that the set

$$U = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \;\middle|\; a, b \in \mathbb{R}, (a, b) \neq (0, 0) \right\}$$

is a subgroup of $\big(\mathrm{Gl}_2(\mathbb{R}), \cdot\big)$.

Exercise 1.42
Let $(G, \cdot)$ be a group, $g \in G$ and $\emptyset \neq U \subseteq G$ a finite subset of G.

a. Is $\{g^n \mid n > 0\}$ finite, then there is an $n > 0$ with $g^n = e_G$.

b. U a subgroup of G if and only if for all $u, v \in U$ also $u \cdot v \in U$.

Exercise 1.43
Which of the following subsets are subgroups of $\big(\mathrm{Sym}(\mathbb{R}), \circ\big)$?

a. $U = \{f \in \mathrm{Sym}(\mathbb{R}) \mid f(x) < f(y) \text{ if } x > y\}$,

b. $V = \{f \in \mathrm{Sym}(\mathbb{R}) \mid |f(x)| = |x| \text{ for all } x \in \mathbb{R}\}$.

Exercise 1.44
For two real numbers $a, b \in \mathbb{R}$ we define the map

$$f_{a,b} : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto a \cdot x + b.$$

Which of the following sets is a subgroup of $\big(\mathrm{Sym}(\mathbb{R}), \circ\big)$?

a. $U = \{f_{a,b} \mid a, b \in \mathbb{R}, a \neq 0\}$,

b. $V = \{f_{a,1} \mid a \in \mathbb{R}, a \neq 0\}$.

Exercise 1.45
Let $(G, \cdot)$ be a group. Show that the set

$$Z(G) := \{g \in G \mid g \cdot h = h \cdot g \;\; \forall\, h \in H\}$$

is a subgroup of G. It is known as the *centre* of G.

C) Group Homomorphisms

Whenever we define a structure on a set we also consider maps which respect this structure. These will be called *morphisms* or *homomorphisms*.

Definition 1.46
Let $(G, \cdot)$ and $(H, *)$ be two groups. A map $\alpha : G \to H$ is a *group homomorphism* (or for short a *homomorphism*), if

$$\alpha(g \cdot h) = \alpha(g) * \alpha(h)$$

for all $g, h \in G$.

Let us consider first some examples.

Example 1.47

a. If $(G, \cdot)$ is a group and $U \leq G$ is a subgroup subgroup then the canonical inclusion $i_U : U \to G$ is a group homomorphism, since $i_U(g \cdot h) = g \cdot h = i_U(g) \cdot i_U(h)$ for $g, h \in U$.

b. Let $a \in \mathbb{R}$ and $m_a : (\mathbb{R}, +) \to (\mathbb{R}, +) : g \mapsto ag$ be the multiplication by a. Then m_a is a group homomorphism since

$$m_a(g + h) = a(g + h) = ag + ah = m_a(g) + m_a(h)$$

for $g, h \in \mathbb{R}$.

c. Let $(G, \cdot)$ be a group and for $g \in G$ define the maps

$$R_g : G \to G : h \mapsto hg \qquad \text{(die “\textit{right translation}”)}$$

and

$$L_g : G \to G : h \mapsto gh. \qquad \text{(die “\textit{left translation}”)}$$

For $g \neq e$ the cancellation rule implies that

$$L_g(g \cdot g) = g^3 \neq g^4 = L_g(g) \cdot L_g(g)$$

and thus L_g is not a group homomorphism. Analogously also R_g is none. However, one sees easily that L_g and R_g are bijective with inverses $L_{g^{-1}}$ resp. $R_{g^{-1}}$.

d. If $(G, \cdot)$ is a group and $g \in G$ we define

$$i_g : G \to G : h \mapsto ghg^{-1} =: h^g.$$

i_g is known as an *inner automorphism* or *conjugation* with g.

Claim: The conjugation is a bijective group homomorphism.

For $h, k \in G$ we have

$$\begin{aligned} i_g(hk) =& g(hk)g^{-1} = g\big(hek\big)g^{-1} = g\big(h\big(g^{-1}g\big)k\big)g^{-1} \\ =& \big(ghg^{-1}\big)\big(gkg^{-1}\big) = i_g(h) \cdot i_g(k). \end{aligned}$$

Hence i_g is a group homomorphism. Moreover, for an arbitrary $h \in G$ the following holds:

$$(i_g \circ i_{g^{-1}})(h) = g\Big(g^{-1}h\big(g^{-1}\big)^{-1}\Big)g^{-1} = \big(gg^{-1}\big)h\big(gg^{-1}\big) = ehe = h = \mathrm{id}_G(h).$$

Hence, $i_g \circ i_{g^{-1}} = \mathrm{id}_G$. Analogously one checks that $i_{g^{-1}} \circ i_g = \mathrm{id}_G$. Thus i_g bijective with inverse $i_{g^{-1}}$.

With the notation of the above example we have obviously that $i_g = R_g \circ L_{g^{-1}}$. Thus the composition of two non-homomorphisms may very well be a homomorphism. The following lemma states that indeed the composition of two homomorphisms is always a homomorphism.

Lemma 1.48

Let $\alpha_1 : (G_1, \cdot) \to (G_2, *)$ *and* $\alpha_2 : (G_2, *) \to (G_3, \times)$ *be group homomorphisms. Then also* $\alpha_2 \circ \alpha_1 : (G_1, \cdot) \to (G_3, \times)$ *is a group homomorphism.*

Proof: Let $g, h \in G_1$. Then

$$\begin{aligned}(\alpha_2 \circ \alpha_1)(g \cdot h) &= \alpha_2\big(\alpha_1(g \cdot h)\big) = \alpha_2\big(\alpha_1(g) * \alpha_1(h)\big) = \alpha_2\big(\alpha_1(g)\big) \times \alpha_2\big(\alpha_1(h)\big) \\ &= (\alpha_2 \circ \alpha_1)(g) \times (\alpha_2 \circ \alpha_1)(h).\end{aligned}$$

□

Definition 1.49

Let $\alpha : (G, \cdot) \to (H, *)$ be a group homomorphism.

a. We call α a *monomorphism*, if α is injective.

b. We call α an *epimorphism*, if α is surjective.

c. We call α an *isomorphism*, if α is bijective.

d. We call α an *endomorphism*, if $(G, \cdot) = (H, *)$.

e. We call α an *automorphism*, if α is a bijective endomorphism.

f. We say the groups $(G, \cdot)$ and $(H, *)$ are *isomorphic* if there exists an isomorphism $\alpha : G \to H$. We then write for short $G \cong H$.

Example 1.50

a. In Example 1.47 m_a is an endomorphism. Moreover, m_a is an automorphism with inverse $m_{\frac{1}{a}}$ if and only if $a \neq 0$.

b. Is $(G, \cdot)$ a group and $g \in G$ then the conjugation i_g by g is an automorphism by Example 1.47, and its inverse is $i_{g^{-1}}$.

c. The map $\det : \big(\mathrm{Gl}_2(\mathbb{R}), \cdot\big) \longrightarrow (\mathbb{R} \setminus \{0\}, \cdot)$ from Exercise 1.21 is an epimorphism.

The fact that a group homomorphism respects the group structure has some simple but nevertheless very important consequences.

Proposition 1.51

Let $\alpha : (G, \cdot) \to (H, *)$ *be a group homomorphism. Then:*

a. $\alpha(e_G) = e_H$.

b. $\alpha\big(g^{-1}\big) = \big(\alpha(g)\big)^{-1}$ *for* $g \in G$.

c. $\alpha\big(g^n\big) = \big(\alpha(g)\big)^n$ *for* $g \in G$ *and* $n \in \mathbb{Z}$.

d. *If* α *is bijective then* $\alpha^{-1} : H \to G$ *is a group homomorphism.*

e. *If* $U \leq G$ *then* $\alpha(U) \leq H$. $\alpha(U)$ *is the* image *of* U *by* α.

f. *If* $V \leq H$ *then* $\alpha^{-1}(V) \leq G$. $\alpha^{-1}(V)$ *is the* preimage *of* V *by* α.

g. $\mathrm{Im}(\alpha) := \alpha(G)$*, the so called* image *of* α*, is a subgroup of* H.

h. $\mathrm{Ker}(\alpha) := \alpha^{-1}(e_H)$*, the so called* kernel *of* α*, is a subgroup of* G.

Proof: a. We have

$$e_H * \alpha(e_G) = \alpha(e_G) = \alpha(e_G \cdot e_G) = \alpha(e_G) * \alpha(e_G).$$

With the aid of the cancellation rule 1.6 we then get $e_H = \alpha(e_G)$.

b. For $g \in G$ we get

$$\alpha\big(g^{-1}\big) * \alpha(g) = \alpha\big(g^{-1} \cdot g\big) = \alpha(e_G) = e_H.$$

The uniqueness of the inverses in H implies the claim.

c. Let $g \in G$ and $n \in \mathbb{Z}$. The case $n \geq 0$ will be proved by induction on n. If $n = 0$ the claim follows by a., and if $n > 0$ by definition and by induction on n we get

$$\alpha(g^n) = \alpha(g^{n-1} \cdot g) = \alpha(g^{n-1}) \cdot \alpha(g) \overset{\text{Ind.}}{=} \alpha(g)^{n-1} \cdot \alpha(g) = \alpha(g)^n. \tag{14}$$

If $n < 0$ then $-n > 0$ and the power laws imply

$$\alpha(g^n) = \alpha\big((g^{-1})^{-n}\big) \overset{(14)}{=} \alpha\big(g^{-1}\big)^{-n} \overset{\text{b.}}{=} \big(\alpha(g)^{-1}\big)^{-n} = \alpha(g)^n.$$

d. If $\alpha : G \to H$ is bijective the inverse mapping $\alpha^{-1} : H \to G$ exists. Let $u, v \in H$. We set $g := \alpha^{-1}(u)$ and $h := \alpha^{-1}(v)$, so that $u = \alpha(g)$ and $v = \alpha(h)$. Then

$$\alpha^{-1}(u * v) = \alpha^{-1}\big(\alpha(g) * \alpha(h)\big) = \alpha^{-1}\big(\alpha(g \cdot h)\big) = g \cdot h = \alpha^{-1}(u) \cdot \alpha^{-1}(v).$$

Hence, α^{-1} is a group homomorphism.

e. Let $u, v \in \alpha(U)$. Then there are $g, h \in U$ such that $\alpha(g) = u$ and $\alpha(h) = v$. Since $g \cdot h \in U$ we get

$$u * v = \alpha(g) * \alpha(h) = \alpha(g \cdot h) \in \alpha(U).$$

Moreover $g^{-1} \in U$ and thus

$$u^{-1} = \big(\alpha(g)\big)^{-1} = \alpha\big(g^{-1}\big) \in \alpha(U).$$

Finally, since $\alpha(e_G) \in \alpha(U)$ we have $\alpha(U) \neq \emptyset$ and Proposition 1.28 implies that $\alpha(U)$ is a subgroup of H.

f. Let $g, h \in \alpha^{-1}(V)$ then $\alpha(g \cdot h) = \alpha(g) * \alpha(h) \in V$, since V is a subgroup. Moreover, $g \cdot h \in \alpha^{-1}(V)$ and $\alpha(g^{-1}) = (\alpha(g))^{-1} \in V$, again since V is a subgroup. Thus g^{-1} belongs to $\alpha^{-1}(V)$. The preimage of V by α is non-empty, since $\alpha(e_G) = e_H \in V$ and therefore $e_G \in \alpha^{-1}(V)$. Applying Proposition 1.28 we get again that $\alpha^{-1}(V)$ is a subgroup of G.

g. This is a special case of e..

h. This is a special case of f..

$\square$

To see whether a map is injective or not we have to check that each element in the image has only a single preimage. We will now see that for group homomorphisms we have to do much less.

Lemma 1.52
A group homomorphism $\alpha : (G, \cdot) \to (H, *)$ *is injective if and only if* $\mathrm{Ker}(\alpha) = \{e_G\}$.

Proof: If α injective then $\alpha^{-1}(e_H)$ contains at most one element. Due to $\alpha(e_G) = e_H$ it certainly contains the element e_G, and thus $\mathrm{Ker}(\alpha) = \alpha^{-1}(e_H) = \{e_G\}$.

Suppose now that $\mathrm{Ker}(\alpha) = \{e_G\}$ and let $g, h \in G$ such that $\alpha(g) = \alpha(h)$. Then

$$e_H = \alpha(g) * (\alpha(h))^{-1} = \alpha(g) * \alpha(h^{-1}) = \alpha(g \cdot h^{-1}).$$

It follows that $g \cdot h^{-1} = e_G$ and hence, $g = h$. This proves that α is injective. $\square$

Exercise 1.53
Consider the group $(G, \cdot)$ in Exercise 1.18 b. and the group $(U, \cdot)$ in Exercise 1.41. Show that the map

$$\alpha : G \longrightarrow U : (a, b) \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

is an isomorphism of groups.

Exercise 1.54
We consider the group $U = \{f_{a,b} : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto ax + b \mid a, b \in \mathbb{R}, a \neq 0\}$ from Exercise 1.44 where the group operation is the composition of maps, and we consider the group $(\mathbb{R} \setminus \{0\}, \cdot)$. Show that the map

$$\alpha : U \longrightarrow \mathbb{R} \setminus \{0\} : f_{a,b} \mapsto a$$

is a group homomorphism.

Exercise 1.55
Let $(G, \cdot)$ be a group and $g \in G$.

a. The map

$$\alpha : \mathbb{Z} \longrightarrow G : n \mapsto g^n$$

is a group homomorphism with image $\mathrm{Im}(\alpha) = \langle g \rangle$.

b. α is injective if and only if $g^k \neq g^l$ for all $k, l \in \mathbb{Z}$ with $k \neq l$.

c. Suppose there are integers $k \neq l$ such that $g^k = g^l$. Then the number

$$n = \min\{m \in \mathbb{N} \mid m > 0, g^m = e_G\}$$

exists and the following holds true:

(i) $\mathrm{Ker}(\alpha) = \{m \in \mathbb{Z} \mid g^m = e\} = n\mathbb{Z}$,

(ii) $\langle g \rangle = \{e_G, g, g^2, \ldots, g^{n-1}\}$, and

(iii) $|\langle g \rangle| = n$.

Exercise 1.56
Let $(G, \cdot)$ be a group and $h, k \in G$ be fixed. Check for each of the following maps which conditions h and k have to satisfy to ensure that the map is a group homomorphism.

a. $\alpha : G \to G : g \mapsto h \cdot g$,

b. $\alpha : G \to G : g \mapsto h \cdot g \cdot h$,

c. $\beta : G \to G : g \mapsto h^{-1} \cdot g \cdot k$,

Exercise 1.57
Let $(G, \cdot)$ be a group. Show that $\alpha : G \longrightarrow G : g \mapsto g^2$ is a group homomorphism if and only if G is abelian.

Exercise 1.58
Let $(G, \cdot)$ be a group. Show that $\mathrm{inv} : G \longrightarrow G : g \mapsto g^{-1}$ is a group homomorphism if and only if G is abelian.

Exercise 1.59
Let $\alpha : (G, \cdot) \longrightarrow (H, *)$ be a group homomorphism, $g \in G$ and $g' \in \mathrm{Ker}(\alpha)$. Show that then $g^{-1} \cdot g' \cdot g \in \mathrm{Ker}(\alpha)$.

Exercise 1.60
Let $(G, \cdot)$ be a group and $g \in G$.

a. The map $L_g : G \longrightarrow G : h \mapsto g \cdot h$ is bijective.

b. The map $\alpha : G \longrightarrow \mathrm{Sym}(G) : g \mapsto L_g$ is a monomorphism.[12]

It is an important principle that a monomorphism respects all good properties of a group respectively of its elements. The order of an element is an example of such a property and the following exercise is an example for this principle.

Exercise 1.61
Let $\alpha : (G, \cdot) \longrightarrow (H, *)$ be a group homomorphism and $g \in G$. We call the cardinality of the subgroup generated by g the *order* of the element of g, and we denote it by $o(g) := |\langle g \rangle|$.

a. If $o(g) < \infty$ then $o(g)$ a is a multiple of $o\big(\alpha(g)\big)$.

[12]This statement is known as the *Theorem of Cayley.* It says that every finite group is isomorphic to a subgroup of some symmetric group.

b. If α injective then $o(g) = o\big(\alpha(g)\big)$.

Exercise 1.62

Since $(\mathbb{R}, +)$ is a group we can apply Exercise 1.17 to see that also $(\mathbb{R}^2, +)$ with the componentwise addition is a group. Show that the following map is a group homomorphism:

$$\alpha : \mathbb{R}^2 \longrightarrow \mathbb{R} : (x, y) \mapsto 2x + 3y$$

Compute the image and the kernel of α. Is α injective or surjective?

Exercise 1.63

Since $(\mathbb{R} \setminus \{0\}, \cdot)$ and $(\mathbb{R}, +)$ are groups by Exercise 1.17 the set $G = (\mathbb{R} \setminus \{0\}) \times \mathbb{R}$ is a group with the operation

$$(r, s) * (r', s') := (r \cdot s', r + s').$$

Moreover, we know that the complex numbers $(\mathbb{C} \setminus \{0\}, \cdot)$ without zero form a group with respect to the multiplication. Show that the map

$$\alpha : G \longrightarrow \mathbb{C} \setminus \{0\} : (r, s) \mapsto r \cdot \exp(s \cdot \pi \cdot i)$$

is a group homomorphism. Compute the image and the kernel of α. Is α injective or surjective? Here, π is the circle constant and i the imaginary unit.

Exercise 1.64

Find all group homomorphisms from $(\mathbb{Z}, +)$ to $(\mathbb{R}, +)$.

2 Equivalence Relations

Equivalence relations are a very important mathematical principle for either *ordering* things or *constructing* new stuff, which we will use in the sequel at several crucial points. We therefore would like to take a minute consider this principle before going on with the theory of groups.

A *relation* on a set M is just a subset of the Cartesian product $M \times M$. The idea is that an element x is somehow related to an element y if the pair (x, y) is contained in R. Here the meaning of "*is related to*" is not further specified. We refer to the literature to see how maps can be considered as special cases of relations and for the notion of order relations, where x is related to y if x is smaller than y in some sense. These to types of relations usually do not create much of a problem for beginners. Unfortunately, the same cannot be said for the equally important *equivalence relations*. The basic principle, however, is very simple and we want to explain it first with an example.

The students in a school are usually divided into forms depending on their age. The headmaster of the school makes sure that each student belongs to some form and only this one form. In a more mathematical formulation the *set* S of all students is divided into *subsets* K_i, called forms, $i = 1, \ldots, k$, which are pairwise disjoint and such that their union is the set of all students. We then say that

$$S = \dot{\bigcup_{i=1}^{k}} K_i$$

is a *partition* of S into the forms $K_1, \ldots, K_k$. For the affiliation of the students Alfred, Ben and Christopher to some form we can remark:

1) Alfred belongs to some form.
2) If Alfred belongs to the same form as Ben, then Ben belongs to the same form as Alfred.
3) If Alfred belongs to the same form as Ben and Ben belongs to the same form as Christopher, then Alfred belongs to the same form as Christopher.

These statements are so obvious that one cannot believe they have any further significance. However, let us suppose for a moment that the headmaster has made a list for each form which contains the names of the students which belong to this form, and let us suppose he has not yet checked whether every student belongs precisely to one form now, and let us finally suppose that the above rules 1)–3) hold true if we replace the names Alfred, Ben and Christopher by arbitrary students, then, we claim, the headmaster can be sure that he has done a good job and that every student indeed is in one form and one form only.

As a mathematician we are looking for simple rules which ensure that we get a partition of a set in pairwise disjoint subsets, and it turns out that the rules 1)–3) are precisely what we are looking for. Let us again formulate this in more mathematical terms. We say that the student x *is related to* the student y if x belongs to the same form as the student y does, and we define the subset

$$R = \{(x, y) \in S \times S \mid x \text{ is in the same form as } y\}$$

of $S \times S$. The rules 1)–3) for students $x, y, z \in S$ can then be formulated as:

- $(x, x) \in R$.
- If $(x, y) \in R$ then also $(y, x) \in R$.
- If $(x, y) \in R$ and $(y, z) \in R$ then also $(x, z) \in R$.

Such a relation is called an *equivalence relation*, and we say that students in the same form are *equivalent*. The form itself will be called an *equivalence class*.

Of course, we will now introduce the notion of equivalence relation for arbitrary sets, and we will use it in the sequel to partition a set into pairwise disjoint subsets. We finally will use the sets in the partition as elements of a new set.

Definition 2.1
Let M be a set. An *equivalence relation* on M is a subset $R \subseteq M \times M$ such that for all $x, y, z \in M$ the following hold true:

R1: $(x, x) \in R$, ("reflexivity")
R2: $(x, y) \in R \;\Rightarrow\; (y, x) \in R$, ("symmetry")
R3: $(x, y), (y, z) \in R \;\Rightarrow\; (x, z) \in R$. ("transitivity")

It is common in mathematics to denote equivalence relations in a different, rather intuitive way, and we next want to introduce this notation.

Notation 2.2
Let M a set and R an equivalence relation on M. We define for $x, y \in M$

$$x \sim y \quad :\Longleftrightarrow \quad (x, y) \in R,$$

and we call "$\sim$" an equivalence relation instead of R.

With this notation we can reformulate the above three axioms in Definition 2.1. For $x, y, z \in M$ we have:

R1: $x \sim x$, ("Reflexivity")
R2: $x \sim y \;\Rightarrow\; y \sim x$, ("Symmetry")
R3: $x \sim y, y \sim z \;\Rightarrow\; x \sim z$. ("Transitivity")

Definition 2.3
Let M a set and $\sim$ an equivalence relation on M. For $x \in M$ the set

$$\overline{x} := \{y \in M \mid y \sim x\}$$

is called the *equivalence class* of x. Each element $y \in \overline{x}$ is a *representative* of the class $\overline{x}$.

$$M/\sim := \left\{\overline{x} \mid x \in M\right\}$$

denotes the set of *equivalence classes modulo the equivalence relation* $\sim$.

Beispiel 2.4
We consider the set $M = \mathbb{R}^2$ of points in the real plane and we denote by $|P|$ the distance of P to the origin $(0,0)$. For two points $P, Q \in M$ we define

$$P \sim Q \quad \Longleftrightarrow \quad |P| = |Q|,$$

i.e. we call two points *equivalent* if their distance to the origin coincides. Then $\sim$ is an equivalence relation.

R1: For $P \in M$ we have $|P| = |P|$. Thus $P \sim P$.
R2: If $P, Q \in M$ with $P \sim Q$ then $|P| = |Q|$ and hence $|Q| = |P|$. Thus $Q \sim P$.
R3: If $P, Q, R \in M$ with $P \sim Q$ and $Q \sim R$ then $|P| = |Q|$ and $|Q| = |R|$. But then also $|P| = |R|$ and thus $P \sim R$.

The equivalence class

$$\overline{P} = \{Q \in M \mid |Q| = |P|\}$$

of $P \in M$ is the circle about the origin of radius $|P|$.

We have already seen one example of an equivalence relation from our daily life. Another well known example is given by the rational numbers! A fraction is nothing else but an equivalence class of pairs of integers, and the operation of cancellation, e.g. $\frac{1}{2} = \frac{2}{4}$, means just that we replace one representative of the equivalence class by another one.

Beispiel 2.5
We can define the rational numbers as equivalence classes of pairs of integers as follows. For $(p, q), (p', q') \in M := \mathbb{Z} \times \big(\mathbb{Z} \setminus \{0\}\big)$ we define

$$(p, q) \sim (p', q') \quad :\Longleftrightarrow \quad pq' = p'q.$$

We want to show that this actually defines an equivalence relation on M. For this let $x = (p, q), x' = (p', q'), x'' = (p'', q'') \in M$ be given:[13]

R1: For the reflexivity we have to show $x \sim x$. But the equality $pq = pq$ implies $x = (p, q) \sim (p, q) = x$.
R2: For the symmetry we start with $x \sim x'$ and have to show that then also $x' \sim x$. From $x \sim x'$ we deduce that $pq' = p'q$ and thus $p'q = pq'$. But the latter means that $x' = (p', q') \sim (p, q) = x$.
R3: For the transitivity we finally start with $x \sim x'$ and $x' \sim x''$, and we want to conclude that also $x \sim x''$. Due to $x \sim x'$ we have $pq' = p'q$, and similarly due

[13]Don't let yourself be deceived by the fact that the elements of M are pairs of numbers! If we tried to write the relation as a subset of $M \times M$, this would give the following clumsy description

$$R = \big\{\big((p, q), (p', q')\big) \in M \times M \;\big|\; pq' = p'q\big\}.$$

Maybe this explains why we prefer the *alternative* description.

to $x' \sim x''$ we get $p'q'' = p''q'$. If we multiply the first equation by q'' and the second equation by q we get

$$pq'q'' = p'qq'' = p'q''q = p''q'q.$$

Since $q' \neq 0$ we can divide both sides of the equation by q' and get:

$$pq'' = p''q.$$

This however implies $x = (p, q) \sim (p'', q'') = x''$.

Thus the three axioms of an equivalence relation are fulfilled.

We then set $\mathbb{Q} := M/\sim$, and for $(p, q) \in M$ we set $\frac{p}{q} := \overline{(p, q)}$, i.e. the rational number $\frac{p}{q}$ is the equivalence class of the pair (p, q). The equivalence class $\sim$ then implies that $\frac{p}{q}$ and $\frac{p'}{q'}$ coincide, if the crosswise products of numerator and denominator, pq' and $p'q$, coincide, or in a more familiar formulation that the fractions coincide once we have *expanded* the fractions by q' respectively by q: $\frac{p}{q} = \frac{pq'}{qq'} \overset{!}{=} \frac{p'q}{q'q} = \frac{p'}{q'}$.

Also the rules for addition and multiplication of rational numbers can easily be formulated with the notion of equivalence classes. For $(p, q), (r, s) \in M$ define:

$$\begin{aligned}\overline{(p, q)} + \overline{(r, s)} &:= \overline{(ps + qr, qs)},\\ \overline{(p, q)} \cdot \overline{(r, s)} &:= \overline{(pr, qs)}.\end{aligned}$$

There is, however, one problem one has to come about with this definition, namely, that it does not depend on the particular representatives that we have chosen. We say that we have to show that the operation is *well defined* (see also Footnote 22 on page 55). We demonstrate this for the addition of rational numbers.

Let $(p', q') \in \overline{(p, q)}$ and $(r', s') \in \overline{(r, s)}$ be possibly other representatives of the classes then $p'q = q'p$ and $r's = s'r$. We have to show that $(p's' + q'r', q's') \in \overline{(ps + qr, qs)}$. This is guaranteed by the following equation:

$$(p's' + q'r')(qs) = p'qs's + q'qr's = q'ps's + q'qs'r = (ps + qr)(q's').$$

$\square$

At the beginning of this section we claimed that the three axioms of an equivalence relation on a set M would guarantee that the equivalence classes induce a partition of M into disjoint subsets. We will now proof this fact, but for this we first define what a partition is.

Definition 2.6

a. Two sets M and N are *disjoint* if $M \cap N = \emptyset$.

b. A family $(M_i)_{i \in I}$ of sets is *pairwise disjoint* if M_i and M_j are disjoint for all $i, j \in I$ with $i \neq j$.

c. Let M be a set. A pairwise disjoint family $(M_i)_{i\in I}$ of subsets of M is called a *partition* of M if $M = \bigcup_{i\in I} M_i$. In this case we write:

$$M = \dot{\bigcup_{i\in I}} M_i.$$

Proposition 2.7

Let $(M_i)_{i\in I}$ be a partition of M and define a relation on M by

$$x \sim y \iff \exists\, i \in I \,:\, x, y \in M_i.$$

Then $\sim$ is an equivalence relation on M.

Proof: If $x \in M = \dot{\bigcup}_{i\in I} M_i$ then there is an $i \in I$ with $x \in M_i$ and thus $x \sim x$. Hence, $\sim$ is reflexive.

Let $x, y \in M$ with $x \sim y$ then there is an $i \in I$ with $x, y \in M_i$. But then we also have $y \sim x$, and the relation is thus symmetric.

Let $x, y, z \in M$ with $x \sim y$ and $y \sim z$ then there are $i, j \in I$ with $x, y \in M_i$ and $y, z \in M_j$. Since the sets in the partition are pairwise disjoint and since $y \in M_i \cap M_j$ we deduce that $M_i = M_j$. Hence, $x, z \in M_i$ and $x \sim z$. Therefore, $\sim$ is also transitive. $\square$

Proposition 2.8

Let M be a set. If $\sim$ is an equivalence relation on M then the equivalence classes form a partition of M, i.e. each element of M belongs to precisely on equivalence class.

In particular, for $x, y \in M$ we have either $\overline{x} = \overline{y}$ or $\overline{x} \cap \overline{y} = \emptyset$.

Proof: Let $x \in M$ be an arbitrary element. From $x \sim x$ we deduce that $x \in \overline{x} \subseteq \bigcup_{\overline{y}\in M/\sim} \overline{y}$. Thus

$$M = \bigcup_{\overline{y}\in M/\sim} \overline{y}.$$

It remains to show that the equivalence classes are pairwise disjoint.

Let $\overline{x}, \overline{y} \in M/\sim$ with $\overline{x} \cap \overline{y} \neq \emptyset$. Then there is a $z \in \overline{x} \cap \overline{y}$, and we have $z \sim x$ and $z \sim y$. By symmetry we also get $x \sim z$ and then $x \sim y$ by transitivity. Let now $u \in \overline{x}$ be any element. Then $u \sim x$ and again by transitivity we have $u \sim y$. Hence, $u \in \overline{y}$ and therefore $\overline{x} \subseteq \overline{y}$. Exchanging the roles of x and y the same argument shows that $\overline{x} = \overline{y}$. $\square$

Corollary 2.9

Let M be a finite set, $\sim$ be an equivalence relation on M, and $M_1, \ldots, M_s$ be the pairwise different equivalence classes of $\sim$. Then:

$$|M| = \sum_{i=1}^{s} |M_i|.$$

Proof: With M also all M_i are finite and thus the statement follows from Proposition 2.8. □

Aufgabe 2.10
Let $M = \{(a_n)_{n\in\mathbb{N}} \mid a_n \in \mathbb{Q}\}$ be the set of all sequences of rational numbers. Show that

$$(a_n)_{n\in\mathbb{N}} \sim (b_n)_{n\in\mathbb{N}} \quad :\Longleftrightarrow \quad \lim_{n\to\infty}(a_n - b_n) = 0$$

defines an equivalence relation on M.

Aufgabe 2.11
We define for two points $(x, y), (x', y') \in \mathbb{R}^2$

$$(x, y) \sim (x', y') \quad :\Longleftrightarrow \quad |x| + |y| = |x'| + |y'|.$$

Show that $\sim$ is an equivalence relation on $\mathbb{R}^2$. Draw the equivalence classes of $(1, 1)$ and $(-2, 3)$ in the plane $\mathbb{R}^2$.

Aufgabe 2.12 (The projective line)
We define for $v = (v_1, v_2), w = (w_1, w_2) \in \mathbb{R}^2 \setminus \{(0, 0)\}$

$$v \sim w \quad \Longleftrightarrow \quad \exists\, \lambda \in \mathbb{R} \setminus \{0\} \,:\, v = \lambda \cdot w$$

where $\lambda \cdot w := (\lambda \cdot w_1, \lambda \cdot w_2)$.

a. Show that $\sim$ is an equivalence relation on $M = \mathbb{R}^2 \setminus \{(0, 0)\}$. It is usual to denote the equivalence class $\overline{(v_1, v_2)}$ of (v_1, v_2) by $(v_1 : v_2)$, and we call the set $M/\sim$ of equivalence classes the *projective line* over $\mathbb{R}$. We denote it by $\mathbb{P}^1_{\mathbb{R}}$.

b. We define on $\mathbb{P}^1_{\mathbb{R}}$ a binary operation by

$$(v_1 : v_2) \cdot (w_1 : w_2) := (v_1 \cdot w_1 - v_2 \cdot w_2 : v_1 \cdot w_2 + v_2 \cdot w_2).$$

Show that this operation is well defined, i.e. it does not depend on the choice of the representative of the equivalence class, and that $\mathbb{P}^1_{\mathbb{R}}$ with this operation is a group. (You may use the results from the proof of Exercise 1.18 b..)

c. Is $(G, \cdot)$ the group from Exercise 1.18 b., then show that the map

$$\alpha : G \longrightarrow \mathbb{P}^1_{\mathbb{R}} : (a, b) \mapsto (a : b)$$

is a group epimorphism with kernel $\mathrm{Ker}(\alpha) = (\mathbb{R} \setminus \{0\}) \times \{0\}$.

d. The set $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ is the circle of radius one whose centre is the origin $(0, 0)$. Show that the map

$$\Phi : S^1 \longrightarrow \mathbb{P}^1_{\mathbb{R}} : (x, y) \mapsto \overline{(x, y)}$$

is surjective.

e. Suppose we allow in the definition of $\sim$ all elements $v, w \in \mathbb{R}^2$, not only the non-zero ones. Is $\sim$ then an equivalence relation on $\mathbb{R}^2$? If so, what is the equivalence class of $(0, 0)$?

Aufgabe 2.13 (The integers)
Let $M = \mathbb{N} \times \mathbb{N}$ and let $m = (a, b) \in M$ and $m' = (a', b') \in M$ be two elements in M. We define

$$m \sim m' \quad \longleftrightarrow \quad a + b' = a' + b.$$

Show that $\sim$ is an equivalence relation and that the map

$$\Phi : \mathbb{Z} \longrightarrow M/\sim : z \mapsto \begin{cases} (z, 0), & \text{if } z \geq 0, \\ (0, -z), & \text{if } z < 0 \end{cases}$$

is bijective.

Aufgabe 2.14
Let M be a set and $\sigma \in \mathrm{Sym}(M)$ a bijective map.

a. By

$$a \sim b \quad \Longleftrightarrow \quad \exists\, m \in \mathbb{Z} \,:\, b = \sigma^m(a)$$

for $a, b \in M$ we define an equivalence relation on the set M.

b. Suppose that $\overline{a}$ for $a \in M$ is a *finite* equivalence class with respect to $\sim$ of cardinal number $|\overline{a}| = n < \infty$.
   (i) The minimum $k = \min\{l > 0 \mid \sigma^l(a) = a\}$ exists.
   (ii) For $q \in \mathbb{Z}$ we have $\sigma^{q \cdot k}(a) = a$.
   (iii) $\overline{a} = \{a, \sigma(a), \ldots, \sigma^{k-1}(a)\}$.
   (iv) $\overline{a}$ contains exactly k elements.

c. Let $M = \{1, \ldots, 7\}$ and $\sigma \in \mathrm{Sym}(M) = \mathbb{S}_7$ be given by the value table

a	1	2	3	4	5	6	7
$\sigma(a)$	3	4	1	7	2	6	5

What are the equivalence classes of M with respect to the above equivalence relation?

3 The Symmetric Group

The symmetric group $\mathrm{Sym}(M)$ of bijective selfmaps of a set M is in a certain sense the *mother* of all groups, since each group is isomorphic to a subgroup of $\mathrm{Sym}(M)$ for some set M.[14] However, for an arbitrary set M $\mathrm{Sym}(M)$ is not all to helpful, since we cannot say much about the structure of the group.

For a finite set M this is completely different. First of all it does not make the slightest difference if we consider $\mathrm{Sym}\left(\{m_1, \ldots, m_n\}\right)$, for an arbitrary set $M = \{m_1, \ldots, m_n\}$ of cardinality n, or if we simply study $\mathbb{S}_n = \mathrm{Sym}\left(\{1, \ldots, n\}\right)$. The two groups are isomorphic and we may thus identify them $\mathbb{S}_n$ is very important for practical reasons. In the lecture Grundlagen der Mathematik the group $\mathbb{S}_n$ will be used in connection with the determinant.

Since the set $\{1, \ldots, n\}$ is finite we can describe the permutation $\sigma \in \mathbb{S}_n$ simply by its *value table.*

Definition 3.1
Is $\sigma \in \mathbb{S}_n$ a *permutation* of the set $\{1, \ldots, n\}$ then we can describe σ by the following scheme

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

respectively

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ \sigma(a_1) & \sigma(a_2) & \ldots & \sigma(a_n) \end{pmatrix},$$

if $a_1, \ldots, a_n$ is any alignment of the numbers $1, \ldots, n$.

Example 3.2
The group $\mathbb{S}_n$ is for $n \geq 3$ not abelian since for the permutations

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \in \mathbb{S}_3$$

we have

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

Note that in the scheme it does not depend in which order the numbers 1 to n occur in the first row. E.g.

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

In order to keep things well-arranged we advise, however, to write the numbers in the first row always in an ascending order.

[14] This is the content of the Theorem of Cayley. See Exercise 1.60.

Remark 3.3

The above representation of a permutation has the nice side effect that we can very easily invert it by simply exchanging the two rows of the scheme. I.e. if

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix} \in \mathbb{S}_n$$

then its inverse σ^{-1} is given by

$$\sigma^{-1} = \begin{pmatrix} \sigma(1) & \sigma(2) & \ldots & \sigma(n) \\ 1 & 2 & \ldots & n \end{pmatrix}.$$

Mathematicians are *lazy* people, or maybe one should better say that they are *efficient*. Therefore, they have thought of a way to represent a permutation in such a way that each of the numbers 1 to n has to be written at most once, not twice. For this we need the notion of a cycle, which permutes k of the numbers $1, \ldots, n$ *cyclically*.

Definition 3.4

a. Let $\{1, \ldots, n\} = \{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_{n-k}\}$, $k \geq 2$, and

$$\sigma = \begin{pmatrix} a_1 & a_2 & \ldots & a_{k-1} & a_k & b_1 & \ldots & b_{n-k} \\ a_2 & a_3 & \ldots & a_k & a_1 & b_1 & \ldots & b_{n-k} \end{pmatrix} \in \mathbb{S}_n,$$

then we call σ a **k-cycle**, and we say that it *cyclically permutes* the numbers $a_1, \ldots, a_k$. Such a map can be represented in a much more efficient way by the one-line scheme:

$$\sigma = (a_1 \ldots a_k). \tag{15}$$

b. A $2-\mathrm{cycle}$ is also called a **transposition**. A transposition $\tau = (i\ j)$ is thus a permutation where only the two numbers i and j are exchanged while all the others are fixed.

c. The neutral element of $\mathbb{S}_n$ is by definition $\mathrm{id}_{\{1,\ldots,n\}}$, and we will simply denote it by id.

Remark 3.5

The interpretation of the notation in Equation (15) is obvious, the first element a_1 is mapped to the second element a_2, the second one is mapped to the third one, and so on, while the last one, namely a_k, is mapped to the first one a_1. This closes the *cycle*. Note here that the cycles $(a_1 \ldots a_k)$, $(a_k\ a_1 \ldots a_{k-1})$, etc. all coincide. In order to avoid this ambiguity we recommend to start a cycle always with the smallest of the numbers $a_1, \ldots, a_k$.

So far we have only introduced k-cycles for $k \geq 2$. We now allow the case $k = 1$, e.g. (1) or (3), and define each 1-cycle (a) (with $a \in \{1, \ldots, n\}$) to be the identity. $\square$

Example 3.6
The permutations

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix} \in \mathbb{S}_4 \quad \text{and} \quad \pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 3 & 2 & 5 \end{pmatrix} \in \mathbb{S}_5$$

are all 3-cycles which permute the numbers $1, 4, 2$ cyclically. In the above notation they are thus described as

$$\sigma = (1\ 4\ 2) \quad \text{and} \quad \pi = (1\ 4\ 2).$$

This shows the disadvantage of the new notation. It does not give any information on the set of definition of the permutation. Different maps are represented by the same scheme. However, we are willing to pay this price for a representation which is both shorter and better organised. Moreover, in applications we usually know very well what the set of definition of a permutation in question is, and anyhow the really important information is which numbers are move. □

The representation of a permutation by the cycle scheme would not be very helpful if it only applied to permutations which are indeed cycles, while all other permutations would have to be represented by the clumsier two-row scheme. We will now show that indeed every permutation can be decomposed as a composition of pairwise disjoint cycles.

Theorem 3.7
Is $\sigma \in \mathbb{S}_n$ a permutation there is a partition

$$\{1, \ldots, n\} = \dot{\bigcup_{i=1}^{t}} \{a_{i1}, \ldots, a_{ik_i}\}$$

such that

$$\sigma = (a_{11} \cdots a_{1k_1}) \circ \ldots \circ (a_{t1} \cdots a_{tk_t}).$$

We call this representation the cycle decomposition *of σ, and we say that the cycles are* pairwise disjoint. *Note also that $k_1 + \ldots + k_t = n$ and that $0 \leq k_i \leq n$ for $i = 1, \ldots, t$.*

Proof: In order to find the cycle decomposition we recall the equivalence relation from Exercise 2.14 on the set $\{1, \ldots, n\}$ given by

$$a \sim b \quad \Longleftrightarrow \quad \exists\, m \in \mathbb{Z} \,:\, b = \sigma^m(a)$$

for $a, b \in \{1, \ldots, n\}$. For $a \in \{1, \ldots, n\}$ the equivalence class of a has the form

$$\overline{a} = \left\{a, \sigma(a), \sigma^2(a), \ldots, \sigma^{k-1}(a)\right\}, \tag{16}$$

where

$$k = \min\{l > 0 \mid \sigma^l(a) = a\} = |\overline{a}|.$$

Due to Proposition 2.8 the equivalence classes of $\sim$ form a partition of $\{1, \ldots, n\}$. We thus can choose integers $a_{11}, \ldots, a_{t1} \in \{1, \ldots, n\}$ such that

$$\{1, \ldots, n\} = \dot{\bigcup_{i=1}^{t}} \overline{a_{i1}}.$$

Set $k_i = |\overline{a_{i1}}|$ and $a_{ij} = \sigma^{j-1}(a_{i1})$ then due to (16) we get

$$\{1, \ldots, n\} = \dot{\bigcup_{i=1}^{t}} \{a_{i1}, a_{i2}, \ldots, a_{ik_i}\}. \tag{17}$$

It remains to show that

$$\sigma = \sigma_1 \circ \cdots \circ \sigma_t$$

where $\sigma_i = (a_{i1} \cdots a_{ik_i})$ is a k_i-cycle. For this let $b \in \{1, \ldots, n\}$ so that $b = a_{ij} = \sigma^{j-1}(a_{i1})$ for some $1 \leq i \leq t$ and some $1 \leq j \leq k_i$. We now apply σ to b and get

$$\sigma(b) = \sigma(a_{ij}) = \sigma^j(a_{i1}) = \left\{ \begin{array}{ll} a_{ij+1}, & \text{if } j < k_i, \\ a_{i1}, & \text{if } j = k_i \end{array} \right\} = \sigma_i(b).$$

Since the decomposition in (17) is disjoint and since b as well as $\sigma_i(b)$ are contained in $\{a_{i1}, \ldots, a_{ik_i}\}$, both b and $\sigma_i(b)$ will be fixed by all σ_l for $l \neq i$, i.e.

$$\big(\sigma_1 \circ \cdots \circ \sigma_t\big)(b) = \sigma_i(b) = \sigma(b).$$

This proves the statement of the theorem. □

Remark 3.8

Note that for two disjoint cycles $\sigma = (a_1 \ \ldots \ a_k), \pi = (b_1 \ \ldots \ b_l) \in \mathbb{S}_n$ we have obviously

$$\sigma \circ \pi = \pi \circ \sigma.$$

Sine for $c \in \{a_1, \ldots, a_k\}$ we have $\sigma(c) \in \{a_1, \ldots, a_k\}$ and thus necessarily $c, \sigma(c) \notin \{b_1, \ldots, b_l\}$ so that

$$(\sigma \circ \pi)(c) = \sigma\big(\pi(c)\big) = \sigma(c) = \pi\big(\sigma(c)\big) = (\pi \circ \sigma)(c). \tag{18}$$

In both cases $c \in \{b_1, \ldots, b_l\}$ and $c \notin \{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_l\}$ we show (18) analogously which proves the above claim.

Moreover, it is obvious that the cycle decomposition of σ is *unique* up the order of the cycles, since the elements of the cycles are cyclically permuted by σ.

Finally, also in its cycle representation it is easy to invert a permutation by simply writing in down from back to front. For this note that for a k-cycle $\sigma = (a_1 \ \ldots \ a_k)$ the inverse is obviously given by

$$\sigma^{-1} = (a_k \ a_{k-1} \ \ldots \ a_2 \ a_1)$$

and is again a k-cycle. But this shows that

$$\pi = (a_{11} \cdots a_{1k_1}) \circ \ldots \circ (a_{t1} \cdots a_{tk_t})$$

has the inverse

$$\pi^{-1} = (a_{tk_t} \cdots a_{t1}) \circ \ldots \circ (a_{1k_1} \cdots a_{11}).$$

$\square$

Example 3.9
The permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 3 & 1 \end{pmatrix} \in \mathbb{S}_5$$

has the cycle decomposition

$$\sigma = (1\;2\;5) \circ (3\;4) = (3\;4) \circ (1\;2\;5). \tag{19}$$

Moreover the inverse of σ is given by

$$\sigma^{-1} = (4\;3) \circ (5\;2\;1) = (1\;5\;2) \circ (3\;4).$$

A natural question is how we found the cycle decomposition in (19). We will try to describe the algorithm in this example by words, and it should be straight forward to generalise this to any other example. We start with the smallest number, 1, and we search for its image under σ, i.e. $\sigma(1) = 2$. This gives the starting part of our first cycle:

$$(1\;2$$

We then consider the image of 2 under σ, i.e. $\sigma(2) = 5$, and we get:

$$(1\;2\;5$$

We continue with the image of 5 under σ, i.e. $\sigma(5) = 1$. However, since this is the first element of our first cycle we simply close the cycle,

$$(1\;2\;5),$$

and start with the smallest number in $\{1, \ldots, 5\}$, which is not yet contained in the first cycle, i.e. 3:

$$(1\;2\;5) \circ (3$$

We then again consider the image of 3 under σ, i.e. $\sigma(3) = 4$, and we continue our second cycle in this way:

$$(1\;2\;5) \circ (3\;4$$

Since we have already used all five elements of the set $\{1, \ldots, 5\}$ we know for sure that $\sigma(4) = 3$. This allows us to close the second cycle:

$$\sigma = (1\;2\;5) \circ (3\;4).$$

As already mentioned, there is no number left in $\{1, \ldots, 5\}$ which has not yet been used in one of the cycles created so far. Thus we are done and have found the cycle decomposition of σ.

The algorithm described above follows immediately from the proof of Theorem 3.7, and we can now write our cycle decomposition now also in the following way

$$\sigma = \Big(1 \;\; \sigma(1) \;\; \sigma^2(1)\Big) \circ \big(3 \;\; \sigma(3)\big),$$

where $\sigma^3(1) = 1$ and $\sigma^2(3) = 3$. □

From now on we will switch between the different ways of representing permutations and we will always use the one which fits our purpose best.

Remark 3.10
For very small values n the group $\mathbb{S}_n$ we can write it down completely, but with increasing n the group $\mathbb{S}_n$ becomes very soon gigantic. $\mathbb{S}_1 = \{\mathrm{id}\}$ and $\mathbb{S}_2 = \{\mathrm{id}, (1\ 2)\}$. $\mathbb{S}_3 = \{\mathrm{id}, (1\ 2), (1\ 3), (2\ 3), (1\ 2\ 3), (1\ 3\ 2)\}$ has already six elements, $\mathbb{S}_4$ even 24 and the number of elements in $\mathbb{S}_{60}$ is in the order of 10^{82}. This number is supposedly close to the number of nucleons in the universe.

Proposition 3.11
$|\mathbb{S}_n| = n! = 1 \cdot 2 \cdot 3 \cdots n.$

Proof: A permutation $\sigma \in \mathbb{S}_n$ is defined by the images $\sigma(1), \ldots, \sigma(n)$ of the numbers $1, \ldots, n$, where each of the numbers $1, \ldots, n$ occurs exactly once among the numbers $\sigma(1), \ldots, \sigma(n)$. We now want to count how many possibilities there are for such a permutation. First we have to fix the number $\sigma(1)$, i.e. the image of 1. For this there are n choices. Once we have fixed $\sigma(1)$, there are only $n-1$ choices for the image $\sigma(2)$ of 2. Then for $\sigma(3)$ there are only $n-2$ choices. Going on like this for $\sigma(i)$ we have $n-i+1$ choices, and finally for $\sigma(n-1)$ there are $n-(n-1)+1 = 2$ choices and for $\sigma(n)$ will be fixed. Altogether we get

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$$

choices for a permutation, which finishes the proof. □

This proof was rather informal and we leave it to the reader to make the *going on like this* more rigorous by formulating it properly with an induction.

Remark 3.12
We now want to study transpositions in more details. Note first that for a transposition $\tau \in \mathbb{S}_n$ we have clearly $\tau^{-1} = \tau$ and $\tau^2 = \mathrm{id}$.

Proposition 3.13
Every permutation in $\mathbb{S}_n$, $n \geq 2$, can be decomposed as a composition of at most n transpositions.

Proof: Is $\sigma = (a_1 \ \ldots \ a_k)$ a k-cycle with $k \geq 2$ then obviously

$$\sigma = (a_1\ a_2) \circ (a_2\ a_3) \circ \ldots \circ (a_{k-2}\ a_{k-1}) \circ (a_{k-1}\ a_k) \tag{20}$$

is a product of $k-1$ transpositions. Is $\mathrm{id} \neq \sigma \in \mathbb{S}_n$ any permutation besides the identity, then σ has a cycle decomposition of the form

$$\sigma = \sigma_1 \circ \ldots \circ \sigma_t$$

where $\sigma_i = (a_{i1} \cdots a_{ik_i})$ is a k_i-cycle, by Theorem 3.7. Since disjoint cycles commute with each other we may assume *without loss of generality*[15] that $k_1 \geq k_2 \geq \ldots \geq k_t$. Moreover, since σ is not the neutral element id we know that $k_1 \geq 2$ and that the number $s = \max\{r \mid 1 \leq r \leq t, k_r \geq 2\}$ is defined. Thus $\sigma_i = \mathrm{id}$ for $i = s+1, \ldots, t$ and

$$\sigma = \sigma_1 \circ \ldots \circ \sigma_s$$

is the product of s cycles. Since σ_i can be written as a product of $k_i - 1$ transpositions we get that σ can be decomposed as a product of

$$(k_1 - 1) + \ldots + (k_s - 1) = (k_1 + \ldots + k_s) - s \leq n - 1$$

transpositions. The claim thus follows for $\sigma \neq \mathrm{id}$. However, since $\mathrm{id} = (1\ 2) \circ (1\ 2)$ is the product of two transpositions and since $n \geq 2$, the proof is finished. □

The proof is *constructive*, since the Equation (20) shows how to decompose a cycle as a product of transpositions and thereby reduces the general problem to computing a cycle decomposition.

Corollary 3.14

Every permutation in $\mathbb{S}_n$, $n \geq 2$, can be written as a product of transpositions of two consecutive integers.

Proof: Due to Proposition 3.13 it suffices to show this for a transposition $(i\ j)$ with $i < j$. However, for this we have obviously

$$\begin{aligned}(i\ j) \quad &= \quad (i\ i+1) \circ (i+1\ i+2) \circ \cdots \circ (j-2\ j-1) \circ (j-1\ j) \circ \\ &\quad \circ (j-2\ j-1) \circ \cdots \circ (i+1\ i+2) \circ (i\ i+1).\end{aligned}$$

□

The representation of a permutation as composition of transpositions in not at all unique. However, it turns out that the parity of the number of permutations needed is independent of the chosen decomposition. This is related to the notion of the sign of a permutation.

Definition 3.15

Let $\sigma \in \mathbb{S}_n$ be given.

[15] "*We may assume without loss of generality*" in principle means that we only consider a *special* case and prove this one case. However, the other cases will follow in the same way. Doing only the special case shows the most important idea of the proof and usually saves a lot of notation. One should, however, be careful when stating that something can be assumed without loss of generality, i.e. one should ensure that the remaining cases indeed do not require new arguments!

a. A pair (i,j) of integers with $1 \leq i,j \leq n$ is called an *error pair* of σ if $i<j$ but $\sigma(i) > \sigma(j)$.

b. We define the *sign* of σ by

$$\operatorname{sgn}(\sigma) = \begin{cases} +1, & \text{if } \sigma \text{ has an even number of error pairs,} \\ -1, & \text{if } \sigma \text{ has an odd number of error pairs.} \end{cases}$$

Example 3.16

A transposition $\tau = (i\ j) \in \mathbb{S}_n$, with $i<j$, has precisely the $2 \cdot (j-i-1)+1$ error pairs

$$(i,i+1),(i,i+2),\ldots,(i,j),(i+1,j),(i+2,j),\ldots,(j-1,j),$$

and thus $\operatorname{sgn}(\tau) = -1$.

The permutation

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

has the error pairs $(1,2)$ and $(3,4)$. Hence, $\operatorname{sgn}(\sigma) = 1$. □

Sometimes it is useful to have the following closed formula for the sign of a permutation. We leave its proof to the reader.

Remark 3.17

For $\sigma \in \mathbb{S}_n$ gilt:

$$\operatorname{sgn}(\sigma) = \prod_{1 \leq i<j \leq n} \frac{\sigma(j)-\sigma(i)}{j-i} = \frac{\sigma(2)-\sigma(1)}{2-1} \cdot \frac{\sigma(3)-\sigma(1)}{3-1} \cdots \frac{\sigma(n)-\sigma(n-1)}{n-(n-1)}.$$

□

Much more important than the formula is the following property of the sign, which makes it easy to compute it in concrete examples.

Theorem 3.18

a. *The map*

$$\operatorname{sgn} : (\mathbb{S}_n, \circ) \longrightarrow \big(\{1,-1\}, \cdot\big)$$

is a group homomorphism, i.e. for $\sigma_1, \sigma_2 \in \mathbb{S}_n$ *we have*

$$\operatorname{sgn}(\sigma_1 \circ \sigma_2) = \operatorname{sgn}(\sigma_1) \cdot \operatorname{sgn}(\sigma_2).$$

b. *Is* $\sigma = \tau_1 \circ \cdots \circ \tau_k \in \mathbb{S}_n$ *a composition of* k *transpositions then*

$$\operatorname{sgn}(\sigma) = (-1)^k.$$

c. *Is* $\sigma \in \mathbb{S}_n$ *then* σ *can either be decomposed as an even number of transpositions or as an odd number.*

Proof: Let $\sigma = \sigma' \circ \tau \in \mathbb{S}_n$ with $\sigma' \in \mathbb{S}_n$ and $\tau = (i\ i+1)$ for a $i \in \{1,\ldots,n-1\}$. Is $(i,i+1)$ an error pair of σ' then τ cancels this one our and σ has one error pair less than σ'. Is conversely $(i, i+1)$ not an error pair of σ' then the composition with τ creates this error pair and σ has one error pair more than σ'. Thus

$$\operatorname{sgn}(\sigma) = -\operatorname{sgn}(\sigma') = \operatorname{sgn}(\sigma') \cdot \operatorname{sgn}(\tau).$$

By Corollary 3.14 every permutation is a product of transpositions of consecutive integers.

Let $\sigma_1 = \tilde{\tau}_1 \circ \cdots \circ \tilde{\tau}_r$ and $\sigma_2 = \tilde{\tau}_{r+1} \circ \cdots \circ \tilde{\tau}_{r+s}$ be given as products of such transpositions of consecutive numbers. By induction on $r+s$ we see that

$$\operatorname{sgn}(\sigma_1 \circ \sigma_2) = (-1)^{r+s} = (-1)^r \cdot (-1)^s = \operatorname{sgn}(\sigma_1) \cdot \operatorname{sgn}(\sigma_2).$$

This proves a., and b. follows by induction on k.

For c. let $\sigma = \tau_1 \circ \cdots \circ \tau_k = \tau_1' \circ \cdots \circ \tau_l'$ with transpositions $\tau_i, \tau_j' \in \mathbb{S}_n$. Then by b.

$$(-1)^k = \operatorname{sgn}(\sigma) = (-1)^l,$$

and thus either k and l are both even or both odd. □

Definition 3.19
$\mathbb{A}_n := \operatorname{Ker}(\operatorname{sgn}) = \{\sigma \in \mathbb{S}_n \mid \operatorname{sgn}(\sigma) = 1\}$ is the *alternating group* of degree n.

Remark 3.20
The kernel of the sign consists of all permutations with positive sign. We call these permutations *even*. By Proposition 1.51 the kernel of sign is a subgroup of $\mathbb{S}_n$.

The set $\{\sigma \in \mathbb{S}_n \mid \operatorname{sgn}(\sigma) = -1\}$ is not a subgroup of $\mathbb{S}_n$. It does not even contain the neutral element id of $\mathbb{S}_n$.

Exercise 3.21
Consider the permutations

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 6 & 3 & 7 & 5 & 1 & 4 \end{pmatrix}, \pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 2 & 7 & 1 & 4 & 5 & 6 \end{pmatrix} \in \mathbb{S}_7.$$

a. Compute $\sigma \circ \pi$, $\pi \circ \sigma$, σ^{-1}, π^{-1}.

b. Compute for each of the permutations in a. the cycle decomposition.

c. Write $\sigma \circ \pi$ as a product of transpositions.

d. Write π^{-1} as a product of transpositions of consecutive integers.

e. Compute for each of the permutations in a. the sign.

Exercise 3.22
Find two subgroups of $\mathbb{S}_4$ of cardinality 4 which are not isomorphic to each other.

Exercise 3.23
Compute all elements of the subgroup $\mathbb{D}_8 = \langle (1\ 2\ 3\ 4), (2\ 4) \rangle \leq \mathbb{S}_4$ of $\mathbb{S}_4$.

Exercise 3.24
Compute all elements of the subgroup $\mathbb{D}_{10} = \langle (1\ 2\ 3\ 4\ 5), (1\ 5) \circ (2\ 4) \rangle \leq \mathbb{S}_5$ of $\mathbb{S}_5$.

Exercise 3.25
Show that a k-cycle σ has the order[16] k and the sign $(-1)^{k-1}$.

Remark 3.26
For $n \in \mathbb{Z}$ with $n \geq 3$ consider the two permutations

$$\pi_n = (1\ 2\ \ldots\ n-1\ n)$$

and

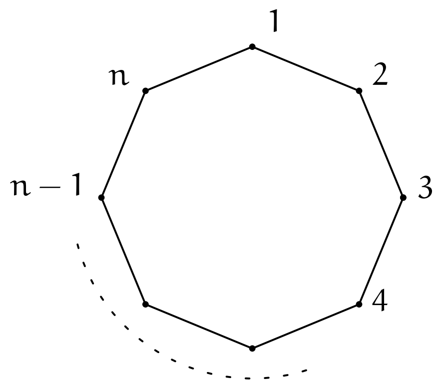
$$\sigma_n = \begin{cases} (1\ n) \circ (2\ n-1) \circ \ldots \circ \left(\frac{n}{2}\ \frac{n}{2}+1\right), & \text{if } n \text{ even,} \\ (1\ n) \circ (2\ n-1) \circ \ldots \circ \left(\frac{n-1}{2}\ \frac{n+3}{2}\right), & \text{if } n \text{ odd} \end{cases}$$

in $\mathbb{S}_n$. They generate the so called *dihedral group*

$$\mathbb{D}_{2n} = \langle \pi_n, \sigma_n \rangle \leq \mathbb{S}_n$$

of order $2n$.

If we label the vertices of a regular n-gon clockwise from 1 to n,



then π_n can be interpreted as a clockwise rotation of the n-gon by the angle $\frac{2\pi}{n}$. It maps the vertex with label 1 to the vertex with label 2, that with label 2 to that with label 3, and so on. Similarly, σ_n can be interpreted as reflection. The dihedral group $\mathbb{D}_{2n}$ is then the full symmetry group of the regular n-gon. Each element corresponds either to a rotation or to a reflection. (See also Example 1.30.)

The groups $\mathbb{D}_8$ and $\mathbb{D}_{10}$ in the Exercises 3.23–4.18 are special cases of such dihedral groups. They are the symmetry groups of the square and the regular pentagon respectively.

[16]For the notion of the order of an element see Exercise 1.61

4 Normal Subgroups and Quotient Groups

The notion *quotient group* is without any doubt one of the most difficult notions for beginners in mathematics. The first examples of groups that we encountered were $(\mathbb{Z},+)$ and $(\mathbb{Q},+)$ – the elements are simply numbers, i.e. something rather familiar. Afterwards we considered the symmetric group whose elements are maps. For an arbitrary set M the group $\mathrm{Sym}(M)$ is certainly not that easy. The special case of the $\mathbb{S}_n$, however, is much simpler again, since we could do calculations with matrix-like schemes and we had fixed rules on how to manipulate these. The step to the quotient group seems to be a major step which requires an awful lot of abstraction, since the elements of these groups turn out to be sets. As mentioned, this leads to a lot of trouble for beginners — however, I would like to convince you that quotient groups are in principle something very simple. If we forget (like with the permutations in $\mathbb{S}_n$) what the elements actually are, there remain only three simple rules for manipulating and computing with the elements of a quotient group. And all you need to remember, to be able to master quotient groups, are these rules!

Like all groups quotient groups consist of a set together with a group operation. The first two subsections of this section are devoted to introducing the underlying set.

A) The Theorem of Lagrange

In this section we consider a particular type of equivalence relation. The underlying set will be a group, and for the definition of the equivalence relation we will use the following notion.

Notation 4.1
Let $(G,\cdot)$ be a group and $A, B \subseteq G$ be two subsets of G. We define

$$A \cdot B = \{a \cdot b \mid a \in A, b \in B\}.$$

Sometimes we write short AB for $A \cdot B$, and if $A = \{g\}$ consists only of one element we write gB instead of $\{g\}B$ and Bg instead of $B\{g\}$.

Note that the associativity of the group operation induces the associativity of the product of subsets, i.e. for $A, B, C \subseteq G$ we have

$$\begin{aligned}(A \cdot B) \cdot C = \{(a \cdot b) \cdot c \mid a \in A, b \in B, c \in C\} \\ = \{a \cdot (b \cdot c) \mid a \in A, b \in B, c \in C\} = A \cdot (B \cdot C).\end{aligned}$$

Proposition 4.2
Let G be a group and $U \leq G$. For two elements $g, h \in G$ we define

$$g \sim h \quad :\Longleftrightarrow \quad g^{-1}h \in U.$$

Then $\sim$ is an equivalence relation on the set G and the equivalence class which corresponds to g is

$$\overline{g} = gU = \{gu \mid u \in U\}.$$

We call gU *the* left coset *of* U *in* G *associated to* g*, and* g *is called a* representative *of the left coset. Moreover we call*

$$G/U = \{gU \mid g \in G\}$$

the set of all left cosets of U *in* G *and we call the cardinality*

$$|G : U| := |G/U|$$

of G/U *the* index *of* U *in* G.

Proof: We have to show that the relation defined by $\sim$ on G is reflexive, symmetric and transitive. For this let $g, h, k \in G$.

R1: Since $g^{-1}g = e \in U$ we have $g \sim g$ and $\sim$ reflexive.
R2: Suppose $g \sim h$ and hence $g^{-1}h \in U$. The closedness of U with respect to inverses implies $h^{-1}g = \left(g^{-1}h\right)^{-1} \in U$. Hence $h \sim g$, and $\sim$ is symmetric.
R3: Suppose $g \sim h$ and $h \sim k$ and hence $g^{-1}h \in U$ and $h^{-1}k \in U$. The closedness of U with respect to the group operation implies $g^{-1}k = \left(g^{-1}h\right)\left(h^{-1}k\right) \in U$ and hence $g \sim k$. $\sim$ is thus also transitive.

Thus $\sim$ is an equivalence relation.

It remains to show that the set elements which are equivalent to g is gU. Is $h \in G$ with $g \sim h$ then by definition $g^{-1}h \in U$ and thus $h = g \cdot (g^{-1}h) \in gU$. Is conversely $h = gu \in gU$ with $u \in U$ then $g^{-1}h = g^{-1}gu = u \in U$ and thus $g \sim h$. $\square$

Since an equivalence relation induces a partition on the underlying set (see Proposition 2.8) we get the following corollary for free.

Corollary 4.3
Let G *be a group and* $U \leq G$*. For* $g, h \in G$ *we have either* $gU = hU$ *or* $gU \cap hU = \emptyset$*, and* G *is disjoint Union of the left cosets of* U *in* G*:*[17]

$$G = \dot{\bigcup_{\lambda \in G/U}} g_\lambda U,$$

where $g_\lambda \in G$ *is any representative of the left coset* λ*, i.e.* $\lambda = g_\lambda U$.

Example 4.4
Consider the group $G = \mathbb{S}_3$ and the subgroup $U = \mathbb{A}_3$. Then there are two cosets:

$$\mathbb{A}_3 = \mathrm{id}\,\mathbb{A}_3 = (1\ 2\ 3)\mathbb{A}_3 = (1\ 3\ 2)\mathbb{A}_3 = \{\mathrm{id}, (1\ 2\ 3), (1\ 3\ 2)\}$$

and

$$(1\ 2)\mathbb{A}_3 = (1\ 3)\mathbb{A}_3 = (2\ 3)\mathbb{A}_3 = \{(1\ 2), (1\ 3), (2\ 3)\}.$$

Thus the index $|\mathbb{S}_3 : \mathbb{A}_3|$ of $\mathbb{A}_3$ in $\mathbb{S}_3$ is two.

[17] Note that in the union $\dot{\bigcup}_{\lambda \in G/U} g_\lambda U$ each coset of U in G occurs *exactly once*, since for each coset we have chosen exactly one representative.

Remark 4.5
One left coset of U in G is always known. It is independent of the concretely given U and G known, namely the left coset whose representative is the neutral element:

$$eU = U$$

i.e. the subgroup itself is always a left coset.

Moreover, one should note that the possible representatives of a left coset are precisely the elements in that coset. In particular, $uU = U$ for each $u \in U$. □

The possibly most important example in our lecture is the set $\mathbb{Z}/n\mathbb{Z}$ of the left cosets of the subgroup $n\mathbb{Z}$ in the group $(\mathbb{Z}, +)$. In order to be able to describe all left cosets in this example and in order to give for each of them a *most simple* representative we need the principle of *division with remainder* for integers.

Proposition 4.6
Is $(G, \cdot) = (\mathbb{Z}, +)$ *and* $U = n\mathbb{Z}$ *for a natural number* $n \geq 1$ *then* U *has precisely* n *left cosets in* G*, namely:*[18]

$$\begin{array}{ccccc} \overline{0} & = & 0 + n\mathbb{Z} & = & n\mathbb{Z}, \\ \overline{1} & = & 1 + n\mathbb{Z} & = & \{1 + nz \mid z \in \mathbb{Z}\}, \\ \overline{2} & = & 2 + n\mathbb{Z} & = & \{2 + nz \mid z \in \mathbb{Z}\}, \\ \vdots & & & & \\ \overline{n-1} & = & (n-1) + n\mathbb{Z} & = & \{n - 1 + nz \mid z \in \mathbb{Z}\}. \end{array}$$

The index $|\mathbb{Z} : n\mathbb{Z}|$ *of* $n\mathbb{Z}$ *in* $\mathbb{Z}$ *is thus* n.

Proof: We have to show that each integer $m \in \mathbb{Z}$ belongs to one of the above mentioned equivalence classes and that they are pairwise different.

Let $m \in \mathbb{Z}$ an arbitrary integer then there exists by division with remainder integers $q, r \in \mathbb{Z}$ such that

$$m = qn + r \quad \text{with} \quad 0 \leq r \leq n - 1.$$

But this implies[19]

$$m - r = qn = nq \in n\mathbb{Z} = U.$$

Thus m is equivalent to r, and therefore $m \in \overline{r}$, where $\overline{r}$ is one of the above n equivalence classes.

It remains to show for $0 \leq i < j \leq n - 1$ that $\overline{i} \neq \overline{j}$. Suppose $\overline{i} = \overline{j}$ then j would be equivalent to i and hence $j - i \in n\mathbb{Z}$ would be a multiple of n. By assumption,

[18] Note here that the group operation is addition, so that a left coset is not denoted by "$g \cdot U$" but by "$g + U$". Maybe this would not be so deceiving if not at the same time the subgroup $U = n\mathbb{Z}$ itself looked like a multiplicative left coset — which it is not! This is one of the main reasons why we prefer the notation $\overline{k}$ instead of $k + n\mathbb{Z}$.

[19] Note again that the group operation in $(\mathbb{Z}, +)$ is the addition. The condition "$g^{-1}h \in U$" translates thus to "$-g + h \in U$". And since the addition is commutative we usually prefer to write "$h - g \in U$".

however, we know that $0 < j - i < n$ is not a multiple of n. This shows that $\overline{i}$ and $\overline{j}$ do not coincide. □

Notation 4.7
From now on we will usually write $\mathbb{Z}_n$ instead of $\mathbb{Z}/n\mathbb{Z}$. Moreover, we sometimes write $\overline{a}_n$ instead of $\overline{a}$ for a coset in $\mathbb{Z}_n$ if we want to indicate modulo which integer we are currently working.

Since the set $\mathbb{Z}_n$ is so important for our lecture we will introduce some common notions.

Remark 4.8
Let $n \in \mathbb{Z}$ be fixed. $x, y \in \mathbb{Z}$ are said to be *congruent modulo* n if

$$x - y \in n\mathbb{Z} = \{nq \mid q \in \mathbb{Z}\}.$$

Congruence is exactly the equivalence relation studied in Proposition 4.6, but instead of the symbol "$\sim$" it is common to write

$$x \equiv y\ (n) \quad \text{or} \quad x \equiv y\ (\mathrm{mod}\ n)$$

in order to indicate that x is congruent to y modulo n. □

We want to close this section with an important theorem, the Theorem of Lagrange. Its main statement is that the order of a subgroup divides the order of the group itself. The following lemma is a central building block of the proof of Lagrange's Theorem.

Lemma 4.9
Let G be a group, $U \leq G$ and $g \in G$. Then the map

$$l_g : U \longrightarrow gU : u \mapsto gu$$

is a bijection. In particular, all left cosets of U in G have the same cardinality $|U|$.

Proof: The cancellation rule applied to $l_g(u) = gu = gu' = l_g(u')$ for $u, u' \in U$ implies that $u = u'$. Hence, is l_g injective. Let conversely $h \in gU$ be a any element in gU then by the definition of gU there is a $u \in U$ such that $h = gu$. But then $h = gu = l_g(u)$, and therefore l_g is surjective. The statement on the cardinality of the left cosets then follows by definition, since two sets have the same cardinality if and only if there is a bijection from one to the other. □

Theorem 4.10 (Theorem of Lagrange)
Let G be a finite group and $U \leq G$ be a subgroup of G. Then

$$|G| = |U| \cdot |G : U|.$$

In particular, $|U|$ and $|G/U| = |G : U|$ are divisors of $|G|$.

Proof: Since G is finite also G/U is finite. Let $G/U = \{\lambda_1, \ldots, \lambda_k\}$ and let the λ_i be pairwise different, in particular $|G : U| = |G/U| = k$. Since the elements of G/U are left cosets of U in G we can for each λ_i chose a representative $g_i \in G$ such that $\lambda_i = g_i U$. By Corollary 2.9 and Lemma 4.9 we then get:

$$|G| \overset{2.9}{=} \sum_{i=1}^{k} |\lambda_i| = \sum_{i=1}^{k} |g_i U| \overset{4.9}{=} \sum_{i=1}^{k} |U| = |U| \cdot k = |U| \cdot |G : U|.$$

□

The Theorem of Lagrange implies immediately the following corollary.

Corollary 4.11

If G is a group and $g \in G$ we define the order *of g as $o(g) := |\langle g \rangle|$, and if G is finite then $o(g)$ is a divisor of $|G|$.*

Remark 4.12

If G is a group and $g \in G$ then Exercise 1.55 implies

$$o(g) = \inf\{k > 0 \mid g^k = e\} \in \mathbb{N} \cup \{\infty\}.$$

We want to demonstrate the usefulness of the Theorem of Lagrange by an example.

Example 4.13

Let $U \leq \mathbb{S}_3$ then $|U| \in \{1, 2, 3, 6\}$ due to the Theorem of Lagrange and since $|\mathbb{S}_3| = 3! = 6$.

1st Case: $|U| = 1$: Necessarily $U = \{\mathrm{id}\}$, since the neutral element of $\mathbb{S}_3$ must be in U.

2nd Case: $|U| = 6$: Since U is a subset of $\mathbb{S}_3$ we must have $U = \mathbb{S}_3$.

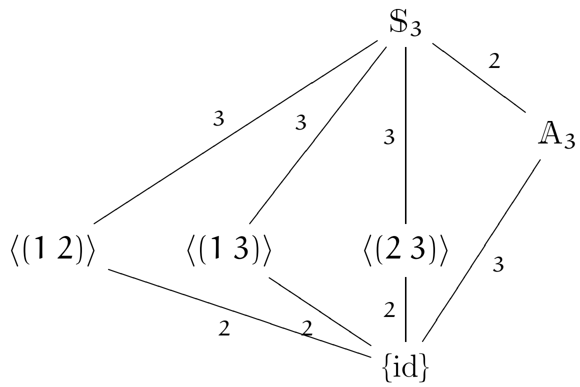
3rd Case: $|U| = 2$: There is some element $\mathrm{id} \neq \sigma \in U$ and thus $o(\sigma) \neq 1$. Corollary 4.11 implies that $o(\sigma)$ is a divisor of $|U| = 2$. Since 2 a prime number we necessarily have $o(\sigma) = 2$ and $U = \langle \sigma \rangle$ is of generated by σ. Hence $\sigma \in \{(1\ 2), (1\ 3), (2\ 3)\}$ and we have three subgroups of order 2:

$$U = \{\mathrm{id}, (1\ 2)\} \text{ or } U = \{\mathrm{id}, (1\ 3)\} \text{ or } U = \{\mathrm{id}, (2\ 3)\}.$$

4th Case: $|U| = 3$: As in the third case there is a $\mathrm{id} \neq \sigma \in U$ and $1 \neq o(\sigma) \mid |U| = 3$. Since also 3 is a prime number we have $o(\sigma) = 3$ and $U = \langle \sigma \rangle$. But then $\sigma \in \{(1\ 2\ 3), (1\ 3\ 2)\}$ and

$$U = \langle (1\ 2\ 3) \rangle = \langle (1\ 3\ 2) \rangle = \{\mathrm{id}, (1\ 2\ 3), (1\ 3\ 2)\} = \mathbb{A}_3.$$

We now know all subgroups of $\mathbb{S}_3$ and can illustrate them in the following *subgroup diagram*:



A line between two groups means that the one further up contains the one further down, and the integers at the lines are the indices of the smaller groups in the larger ones.

Remark 4.14

We called the equivalence classes with respect to the equivalence relation introduced in Proposition 4.2 the *left* cosets, since they had the form gU, i.e. we constructed them by multiplying the subgroup U from the left hand side by some element g. Analogously we could consider the relation:

$$g \sim h \iff hg^{-1} \in U.$$

This leads in the same way to an equivalence relation, and the equivalence class corresponding to g is Ug. It is called a *right coset*. The analogon of Lemma 4.9 for right cosets holds as well, and thus so does the corresponding version of the Theorem of Lagrange, namely, in a finite group G with subgroup U which has m distinct right cosets we have $|G| = |U| \cdot m$. In particular, the number m of right cosets of U in G and the number $|G : U|$ of left cosets of U in G must coincide!

However, in general it is not true that left and right cosets coincide, i.e. it is not in general true that $gU = Ug$. For this consider the example $G = \mathbb{S}_3$, $U = \{\mathrm{id}, (1\ 2)\}$ and $g = (1\ 3)$. An easy computation then shows that $gU \neq Ug$. In the following subsection we want to study subgroups for which the relation $gU = Ug$ always holds true. □

Exercise 4.15 (Product formula)

Let $U, V \leq G$ be subgroups of the group $(G, \cdot)$.

a. Show that by
$$(u, v) \sim (u', v') \iff u \cdot v = u' \cdot v'$$
an equivalence relation on the set $U \times V$ is defined.

b. Show that the equivalence class of $(u, v) \in U \times V$ has the form
$$\overline{(u, v)} = \left\{ \left(u \cdot y, y^{-1} \cdot v\right) \mid y \in U \cap V \right\},$$
and that it has the cardinality $|U \cap V|$.

c. If U and V are finite, then the product formula

$$|\mathsf{U}\cdot\mathsf{V}| = \frac{|\mathsf{U}|\cdot|\mathsf{V}|}{|\mathsf{U}\cap\mathsf{V}|}$$

holds true.

Remark 4.16
The formula in the above exercise is particularly useful when the set $\mathsf{U}\cdot\mathsf{V}$ is indeed a subgroup of G. This is, however, not always the case as we can easily deduce from the Theorem of Lagrange: the product of the subgroups $\langle(1\ 2)\rangle$ and $\langle(1\ 3)\rangle$ of $\mathbb{S}_3$ is due to the exercise a subset of cardinality 4 and by Lagrange's Theorem it thus cannot be a subgroup of $\mathbb{S}_3$. We will see in the following section a condition on V which ensures that $\mathsf{U}\cdot\mathsf{V}$ is a subgroup of G (see Lemma 4.29).

Exercise 4.17
If $(\mathsf{G},\cdot)$ is a group and $|\mathsf{G}|$ is a prime number then G is cyclic.

Exercise 4.18
Find all subgroups of the group $\mathbb{D}_8 = \langle(1\,2\,3\,4),(2\,4)\rangle$.

Exercise 4.19
Find all subgroups of the group $\mathbb{D}_{10} = \langle(1\,2\,3\,4\,5),(1\,5)\circ(2\,4)\rangle$.

B) Normal Subgroups

In the section on equivalence relations we have seen that an equivalence relation is the proper method to partition a set. If the set G that we have partitioned had a structure, e.g. was a group, we may very well ask if there is a natural way to pass this structure on to the set of equivalence classes. Concretely, if G is a group and U is a subgroup, is there a *natural* way to define a group operation on G/U?

The *natural* should mean that the definition is somehow obvious. Given two cosets $g\mathsf{U}$ and $h\mathsf{U}$ we would want to define their product. There is of course a natural way to do this; both are subsets of G and we know already how to define the product of such subsets. What we do not know yet is if this product gives a coset again. If think for a second we might see another obvious way to define the product of $g\mathsf{U}$ and $h\mathsf{U}$, namely as $(gh)\cdot\mathsf{U}$. The advantage of the latter definition is that it certainly gives a coset again. However, what we cannot be sure of is that the definition is independent of the chosen representative. In an ideal world these two obvious ways of defining the product of cosets would coincide, solving both of our problems since the first definition is obviously independent of the chosen representative and the second one is obviously a coset. This leads to the notion of normal subgroups.

Definition 4.20
A subgroup $\mathsf{U}\leq\mathsf{G}$ of G is called *normal* or *normal subgroup* if for all $g\in\mathsf{G}$ and $u\in\mathsf{U}$ we have

$$gug^{-1}\in\mathsf{U}. \tag{21}$$

We denote this by $\mathsf{U}\trianglelefteq\mathsf{G}$.

Remark 4.21
In order to show that a *subset* $U \subseteq G$ is a normal subgroup it suffices *not* to check the property (21) for all $g \in G$ and $u \in U$. First one has to make sure that U is a *subgroup* of G! This is an important part of the definition of a normal subgroup and it is often forgotten by beginners.

Example 4.22
If G is a group then the subgroups $\{e\}$ and G are always normal subgroup. We call them *trivial* normal subgroups.

Lemma 4.23
If G is an abelian group then every subgroup of G is a normal subgroup.

Proof: For $g \in G$ and $u \in U \leq G$ we have $gug^{-1} = gg^{-1}u = eu = u \in U$. □

This lemma implies the following example.

Example 4.24
For each $n \in \mathbb{Z}$ the subgroup $n\mathbb{Z}$ of $(\mathbb{Z}, +)$ is a normal subgroup.

Proposition 4.25
Let G be a group and $U \leq G$ be a subgroup. The following statements are equivalent:[20]

a. $U \trianglelefteq G$ *is a normal subgroup of* G.
b. $gUg^{-1} = U$ *for all* $g \in G$.
c. $gU = Ug$ *for all* $g \in G$.
d. $(gU) \cdot (hU) = ghU$ *for all* $g, h \in G$.

Proof: <u>a. $\Longrightarrow$ b.:</u> By the assumption we have $g \cdot U \cdot g^{-1} \subseteq U$ for any $g \in G$. Let's now fix an arbitrary $g \in G$ and apply this inclusion to g^{-1}. We then get

$$g^{-1} \cdot U \cdot \left(g^{-1}\right)^{-1} \subseteq U,$$

and thus

$$U = e \cdot U \cdot e = g \cdot g^{-1} \cdot U \cdot \left(g^{-1}\right)^{-1} \cdot g^{-1} \subseteq g \cdot U \cdot g^{-1} \subseteq U.$$

This, however, implies $g \cdot U \cdot g^{-1} = U$.

<u>b. $\Longrightarrow$ c.:</u> Multiplying the equation $g \cdot U \cdot g^{-1} = U$ by g on the desired equality.

<u>c. $\Longrightarrow$ d.:</u> Note that $U \cdot U = \{n_1 \cdot n_2 \mid n_1, n_2 \in U\} = U$, since $e \in U$! We thus get for $g, h \in G$

$$(gU) \cdot (hU) = (Ug) \cdot (hU) = U \cdot (gh) \cdot U = (gh) \cdot U \cdot U = ghU.$$

[20] In order to show the equivalence of several statement we can do a so called ring closure. It suffices to show that "a. ⇒ b. ⇒ c. ⇒ d. ⇒ a.", since "a. ⇒ b." and "b. ⇒ c." implies e.g. that "a. ⇒ c.", i.e. the seemingly missing implications follow as well.

<u>d. $\Longrightarrow$ a.:</u> Let $g \in G$ and $n \in U$ be given, then

$$g \cdot n \cdot g^{-1} = g \cdot n \cdot g^{-1} \cdot e \in gU \cdot g^{-1}U = g \cdot g^{-1} \cdot U = e_G \cdot U = U.$$

$\square$

Example 4.26
The subgroup $U := \{\mathrm{id}, (1\ \ 2)\} \subset \mathbb{S}_3$ is not a normal subgroup the $\mathbb{S}_3$ since for $\sigma = (2\ 3) \in \mathbb{S}_3$ we have

$$\sigma \circ (1\ 2) \circ \sigma^{-1} = (2\ 3) \circ (1\ 2) \circ (2\ 3) = (1\ 3) \not\in U.$$

A good source to find normal subgroups are group homomorphisms.

Proposition 4.27
If $\alpha : G \to H$ *is a group homomorphism then* $\mathrm{Ker}(\alpha) \trianglelefteq G$ *is a normal subgroup of* G.

Proof: We know by Proposition 1.51 that $\mathrm{Ker}(\alpha) \leq G$ a subgroup of G. Let $u \in \mathrm{Ker}(\alpha)$ and $g \in G$ then

$$\alpha\big(gug^{-1}\big) = \alpha(g) \cdot \alpha(u) \cdot \alpha\big(g^{-1}\big) \overset{u \in \mathrm{Ker}(\alpha)}{=}$$
$$\alpha(g) \cdot e_H \cdot \alpha\big(g^{-1}\big) = \alpha(g) \cdot \alpha\big(g^{-1}\big) = \alpha\big(gg^{-1}\big) = \alpha(e_G) = e_H.$$

Thus $gug^{-1} \in \mathrm{Ker}(\alpha)$ and $\mathrm{Ker}(\alpha) \trianglelefteq G$. $\square$

Example 4.28
Consider the surjective group homomorphism (see Remark 3.20)

$$\mathrm{sgn} : \mathbb{S}_n \to \{-1, 1\}$$

then $\mathrm{Ker}(\mathrm{sgn}) = \mathbb{A}_n$ is a normal subgroup of $\mathbb{S}_n$.

In general the product of two subgroups is not a subgroup any more. However, if one of the two subgroups is a normal subgroup their product will always be a subgroup again.

Lemma 4.29
Let G *be a group,* $U \leq G$ *and* $N \trianglelefteq G$. *Then*

a. $UN \leq G$.

b. $N \trianglelefteq UN$.

c. $U \cap N \trianglelefteq U$.

Proof: Since N is a normal subgroup is by Proposition we have 4.25

$$(UN) \cdot (UN) = U \cdot (NU) \cdot N = U \cdot (UN) \cdot N = (UU) \cdot (NN) = UN,$$

since $UU = U$ and $NN = N$. This in particular shows that $g \cdot h \in UN$ for all $g, h \in UN$.

Let $g = un \in UN$ with $u \in U$ and $n \in N$ then

$$g^{-1} = n^{-1}u^{-1} \in NU \overset{\text{Prop. 4.25}}{=} UN.$$

Since moreover $e = e \cdot e \in UN$ and thus UN is non-empty we have that UN is a subgroup.

This proves part a., and we leave the rest as an exercise for the reader. $\square$

Example 4.30
We consider the subgroups $U = \langle (1\ 2) \rangle$ and $V = \langle (2\ 3) \rangle$ of $\mathbb{S}_3$. Then

$$U \cdot U' = \{\mathrm{id}, (1\ 2), (2\ 3), (1\ 2\ 3)\}$$

and by the Theorem of Lagrange this product cannot be a subgroup of $\mathbb{S}_3$. This shows that the condition $N \trianglelefteq G$ in Lemma 4.29 is essential.

Exercise 4.31
Proof Part b. and c. of 4.29.

Exercise 4.32
Let G be a finite group and $U \leq G$ be a subgroup of index $|G : U| = 2$. Show that U is a normal subgroup of G.

Exercise 4.33
Let G be a group and $N \leq G$ be the unique subgroup of G of order $|N| = n$. Show that $N \trianglelefteq G$ is a normal subgroup.

Exercise 4.34
Find all normal subgroups of the group $\mathbb{D}_8 = \langle (1\ 2\ 3\ 4), (2\ 4) \rangle$.

Exercise 4.35
Find all normal subgroups of the group $\mathbb{D}_{10} = \langle (1\ 2\ 3\ 4\ ), (1\ 5) \circ (2\ 4) \rangle$.

C) **Quotient Group**

We have now gathered all results which we need in order to formulate the theorem on the quotient group. In the hope that the notation $\overline{g}$ for the left coset gU of a subgroup will make it easier to concentrate on the calculations with the elements of the quotient group by simply clouding the fact that the *element* $\overline{g} = gU$ is actually a *set*, we will adopt this notation right away.

Theorem 4.36
Let $(G, \cdot)$ *be a group and* $U \trianglelefteq G$ *be a normal subgroup of* G. *Then*[21]

$$\overline{g} \cdot \overline{h} = \overline{g \cdot h}, \quad \text{for } \overline{g}, \overline{h} \in G/U. \tag{22}$$

With this multiplication as binary operation the set G/U *is a group. The* neutral element *of* $(G/U, \cdot)$ *is the left coset* $\overline{e} = U$*, and the* inverse *of* $\overline{g} = gU \in G/U$ *is the left coset* $\overline{g^{-1}} = g^{-1}U$.

[21] Here the product $\overline{g} \cdot \overline{h} = gU \cdot hU$ is simply the product of subsets of of G as introduced in Notation 4.1.

Moreover, the residue class map

$$\pi : G \to G/U : g \mapsto \overline{g}$$

is an epimorphism of groups with $\mathrm{Ker}(\pi) = U$.

We call G/U *the* quotient group *of* G *by* U.

Proof: The Equality in (22) follows from 4.25 since U is a normal subgroup:

$$\overline{g} \cdot \overline{h} = gU \cdot hU = ghU = \overline{gh}.$$

Let us now show that G/U with this operation is a group.

For $\overline{g}, \overline{h}, \overline{k} \in G/U$ follows by the associativity of the multiplication in G:

$$\left(\overline{g} \cdot \overline{h}\right) \cdot \overline{k} = \overline{gh} \cdot \overline{k} = \overline{(gh)k} = \overline{g(hk)} = \overline{g} \cdot \overline{hk} = \overline{g} \cdot \left(\overline{h} \cdot \overline{k}\right).$$

Moreover, $\overline{e} \cdot \overline{g} = \overline{eg} = \overline{g}$, so that $\overline{e}$ the neutral element of G/U. And

$$\overline{g^{-1}} \cdot \overline{g} = \overline{g^{-1} \cdot g} = \overline{e},$$

so that $\overline{g}$ has the inverse $\overline{g}^{-1} = \overline{g^{-1}}$. Hence G/U is a group and the claims on the neutral and inverse elements hold true.

From the definition of π we get right away

$$\pi(gh) = \overline{gh} \stackrel{(22)}{=} \overline{g} \cdot \overline{h} = \pi(g) \cdot \pi(h)$$

and

$$\mathrm{Ker}(\pi) = \{g \in G \mid \pi(g) = \overline{e}\} = \{g \in G \mid \overline{g} = \overline{e}\} = \overline{e} = U,$$

so that π is a group homomorphism with $\mathrm{Ker}(\pi) = U$. □

Remark 4.37

a. In the proof of the theorem we used that the product $\overline{g} \cdot \overline{h}$ is a product of subsets of G. We will now try very hard to *forget* this particular fact again! We simply remember that each element $\overline{g}$ of G/U is given by some representative g and that all operations are done using these representatives where the operations obey the following simple but important rules:
   (i) $\overline{g} \cdot \overline{h} = \overline{g \cdot h}$,
   (ii) $\overline{g}^{-1} = \overline{g^{-1}}$,
   (iii) $e_{G/U} = \overline{e} = \overline{u}$, whenever $u \in U$.
b. If the group $(G, \cdot)$ is abelian and $U \trianglelefteq G$ then so is $(G/U, \cdot)$, since

$$\overline{g} \cdot \overline{h} = \overline{gh} = \overline{hg} = \overline{h} \cdot \overline{g}.$$

As an immediate consequence we get the special case $\mathbb{Z}_n$ since $n\mathbb{Z}$ is a normal subgroup of $(\mathbb{Z}, +)$.

Corollary 4.38

For $n \in \mathbb{Z}$ *the pair* $(\mathbb{Z}_n, +)$ *is an abelian group, where* $\overline{x} + \overline{y} = \overline{x + y}$ *for* $x, y \in \mathbb{Z}$.

Remark 4.39
The calculations in $\mathbb{Z}_n$ for some well-chosen n is very familiar to us. If our clock strikes nine then we know that in five hours time it is two o'clock. We do our calculations in $\mathbb{Z}_{12}$:

$$\overline{9}+\overline{5}=\overline{14}=\overline{2+1\cdot 12}=\overline{2}.$$

If your watch uses the 24-hour rhythm then you do your computations in $\mathbb{Z}_{24}$. If we have now nine o'clock then 55 hours ago it was two o'clock:

$$\overline{9}-\overline{55}=\overline{-46}=\overline{2-2\cdot 24}=\overline{2}.$$

If we number the week days from one (Monday) to seven (Sunday) then the question which weekday we have in 51 days if today is Monday comes down to a computation in $\mathbb{Z}_7$:

$$\overline{1}+\overline{51}=\overline{52}=\overline{2+7\cdot 7}=\overline{2}.$$

In 51 days it is Tuesday.

In order to get used to computations in $\mathbb{Z}_n$ we recommend to do some computations with times and weekdays considering this new point of view. □

Example 4.40
For groups of small order, i.e. with a small number of elements, it is sensible to write down an addition respectively multiplication table, which shows for every two elements what their product is. In the case of $\mathbb{Z}_n$ we get for $n=2,3,4$ the following tables:

$$\begin{array}{c|cc} + & \overline{0} & \overline{1} \\ \hline \overline{0} & \overline{0} & \overline{1} \\ \overline{1} & \overline{1} & \overline{0} \end{array} \qquad \begin{array}{c|ccc} + & \overline{0} & \overline{1} & \overline{2} \\ \hline \overline{0} & \overline{0} & \overline{1} & \overline{2} \\ \overline{1} & \overline{1} & \overline{2} & \overline{0} \\ \overline{2} & \overline{2} & \overline{0} & \overline{1} \end{array} \qquad \begin{array}{c|cccc} + & \overline{0} & \overline{1} & \overline{2} & \overline{3} \\ \hline \overline{0} & \overline{0} & \overline{1} & \overline{2} & \overline{3} \\ \overline{1} & \overline{1} & \overline{2} & \overline{3} & \overline{0} \\ \overline{2} & \overline{2} & \overline{3} & \overline{0} & \overline{1} \\ \overline{3} & \overline{3} & \overline{0} & \overline{1} & \overline{2} \end{array}$$

□

If for a normal subgroup $N \trianglelefteq G$ the set of left cosets G/N is a group we may ask the question if the knowledge of the subgroups of G gives some information on the subgroups of G/N. And indeed there is a natural one-to-one correspondence of the subgroups of G which contain N and the subgroups of G/N. This one-to-one correspondence respects normal subgroups.

Remark 4.41
Let $(G,\cdot)$ be a group and $N \trianglelefteq G$ be a normal subgroup the following maps are *bijective*:

$$\big\{U \leq G \;\big|\; N \subseteq U\big\} \longrightarrow \big\{\overline{U} \;\big|\; \overline{U} \leq G/N\big\} : U \mapsto U/N$$

and

$$\big\{M \trianglelefteq G \;\big|\; N \subseteq M\big\} \longrightarrow \big\{\overline{M} \;\big|\; \overline{M} \trianglelefteq G/N\big\} : M \mapsto M/N.$$

The proof is left as an exercise for the reader.

From Proposition 1.39 we know that the subgroup of $(\mathbb{Z},+)$ have the Form $m\mathbb{Z}$ for a non-negative integer m and from Example 1.30 we know that $m\mathbb{Z}$ is contained in $n\mathbb{Z}$ if and only if n is a divisor of m. This gives immediately the following corollary.

Corollary 4.42
If $n \in \mathbb{Z}_{>0}$ is a positive integer then

$$\overline{U} \leq \mathbb{Z}_n \iff \exists\, m \in \{1,\ldots,n\} \text{ with } m \text{ divides } n \,:\, \overline{U} = m\mathbb{Z}/n\mathbb{Z} = \langle \overline{m}_n \rangle.$$

In particular each subgroup of $\mathbb{Z}_n$ is cyclic.

Proof: Due to Remark 4.41 it suffices to find the subgroups U of $\mathbb{Z}$ with $n\mathbb{Z} \subseteq U$. By Proposition 1.39 such a subgroup has the form $U = m\mathbb{Z}$ for an integer $m \geq 0$. The Condition $n\mathbb{Z} \subseteq m\mathbb{Z}$ induces by Example 1.30 that n is a multiple of m, i.e. m must lie between 1 and n and m is a divisor of n. $\square$

We next want to compute the order of an element $\overline{m} \in \mathbb{Z}_n$ for positive integers m and n. For this we introduce the following notation.

Notation 4.43
For two integers $a, b \in \mathbb{Z}$ let

$$\operatorname{lcm}(a,b) := \begin{cases} \min\{z > 0 \mid a \text{ and } b \text{ divides } z\}, & \text{if } a, b \neq 0, \\ 0, & \text{if } a = 0 \text{ or } b = 0. \end{cases}$$

We will see later (see Exercise 7.11) that $\operatorname{lcm}(a,b)$ is a *lowest common multiple* of a and b in the sense of Definition 7.4.

Corollary 4.44
Let $m, n \in \mathbb{Z}_{>0}$. Then

$$o(\overline{m}) = \frac{\operatorname{lcm}(m,n)}{m}$$

is the order of $\overline{m} \in \mathbb{Z}_n$.

Proof: Since $(\mathbb{Z}_n,+)$ is a finite additive group the expression for the computation of the order of $\overline{m}$ in Remark 4.12 takes the following form:

$$\begin{aligned} o(\overline{m}) &= \min\left\{k > 0 \,\middle|\, k \cdot \overline{m} = \overline{0}\right\} \\ &= \min\left\{k > 0 \,\middle|\, n \text{ divides } k \cdot m\right\} \\ &= \frac{m \cdot \min\{k > 0 \mid n \text{ divides } k \cdot m\}}{m} \\ &= \frac{\min\left\{m \cdot k \,\middle|\, k > 0, n \text{ divides } k \cdot m\right\}}{m} \\ &= \frac{\min\left\{l > 0 \,\middle|\, n \text{ and } m \text{ divide } l\right\}}{m} \\ &= \frac{\operatorname{lcm}(m,n)}{m}. \end{aligned}$$

$\square$

Exercise 4.45
Prove Remark 4.41.

Exercise 4.46
Find all subgroups of $(\mathbb{Z}_{33}, +)$.

Exercise 4.47
Consider for $m, n, a, b \in \mathbb{Z}_{>0}$ the element $\left(\overline{a}_m, \overline{b}_n\right) \in \mathbb{Z}_m \times \mathbb{Z}_n$ in the group $(\mathbb{Z}_m \times \mathbb{Z}_n, +)$. Show that the order of this element can be computed as follows

$$o\Big(\left(\overline{a}_m, \overline{b}_n\right)\Big) = \operatorname{lcm}\Big(o\left(\overline{a}_m\right), o\left(\overline{b}_n\right)\Big) = \operatorname{lcm}\left(\frac{\operatorname{lcm}(a,m)}{a}, \frac{\operatorname{lcm}(b,n)}{b}\right)$$

and that it is a divisor of $\operatorname{lcm}(m,n)$. In particular, the group $\mathbb{Z}_m \times \mathbb{Z}_n$ is not cyclic if m and n share a common divisor apart from one..

Exercise 4.48
Compute the order of $\left(\overline{6}_{21}, \overline{9}_{33}\right) \in \mathbb{Z}_{21} \times \mathbb{Z}_{33}$.

Exercise 4.49
Let σ and π be two disjoint cycles in $\mathbb{S}_n$ of length k respectively l. Show that $o(\sigma \circ \pi) = \operatorname{lcm}(k,l)$.

D) The Homomorphism Theorem

If there exists a group isomorphism from one group G to another group H then these two groups are the same from the point of view of group theory. All group theoretically interesting properties are preserved by group isomorphism so that it is not necessary to distinguish between isomorphic groups. Thus, if you want to study a certain group and you do not like the way it represents itself to you then you might as well switch to an isomorphic group whose representation you like better. It is therefore interesting to get to know some basic means of showing that certain groups are indeed isomorphic.

Theorem 4.50 (Homomorphism Theorem)
If $\alpha : G \rightarrow H$ is a group homomorphism then the induced map

$$\widetilde{\alpha} : G/\operatorname{Ker}(\alpha) \longrightarrow \operatorname{Im}(\alpha) : \overline{g} \mapsto \alpha(g)$$

is well-defined[22] *and an isomorphism. In particular*

$$G/\operatorname{Ker}(\alpha) \cong \operatorname{Im}(\alpha).$$

[22]The notion *well defined* means in principle just that the definition is a definition at all. What might be the problem? The elements of $G/\operatorname{Ker}(\alpha)$ are by definition left cosets (even though we still try very hard to forget this fact), and as such they are given by representatives which we use for the computations as well as for the definition of the map $\widetilde{\alpha}$. However, in general each coset has many different representatives and in a definition as above it is a priori not at all clear why the allocation $\overline{g} \mapsto \alpha(g)$ does not depend on the choice of the given representative. I.e. if h a another representative of the same left coset, that is if $\overline{g} = \overline{h}$, is it then true that also $\alpha(g) = \alpha(h)$? If this is not the case then we made a bad mistake since in our *definition* we did not specify which representative of a coset one should use! Thus, for the *well definedness* of the map we have to check

Proof: We show first that $\widetilde{\alpha}$ well defined. For this let $\overline{g} = \overline{h} \in G/\operatorname{Ker}(\alpha)$ be given. Then $g^{-1}h \in \operatorname{Ker}(\alpha)$ and thus

$$e_H = \alpha\big(g^{-1}h\big) = \alpha\big(g^{-1}\big)\alpha(h) = \big(\alpha(g)\big)^{-1}\alpha(h).$$

Therefore, $\alpha(g) = \alpha(h)$ and $\widetilde{\alpha}$ is hence well defined.

For $\overline{g}, \overline{h} \in G/\operatorname{Ker}(\alpha)$ we have moreover

$$\widetilde{\alpha}\big(\overline{g} \cdot \overline{h}\big) = \widetilde{\alpha}\big(\overline{gh}\big) = \alpha(gh) = \alpha(g)\alpha(h) = \widetilde{\alpha}\big(\overline{g}\big) \cdot \widetilde{\alpha}\big(\overline{h}\big).$$

Therefore, $\widetilde{\alpha}$ is also a group homomorphism.

$\widetilde{\alpha}$ is obviously surjective. It thus remains to show that $\widetilde{\alpha}$ injective. For this let $\overline{g}, \overline{h} \in G/\operatorname{Ker}(\alpha)$ with $\alpha(g) = \widetilde{\alpha}\big(\overline{g}\big) = \widetilde{\alpha}\big(\overline{h}\big) = \alpha(h)$, then

$$e_H = \big(\alpha(g)\big)^{-1}\alpha(h) = \alpha\big(g^{-1}\big)\alpha(h) = \alpha\big(g^{-1}h\big).$$

I.e. $g^{-1}h \in \operatorname{Ker}(\alpha)$ and hence $\overline{g} = \overline{h}$. This shows that $\widetilde{\alpha}$ is injective. $\square$

Example 4.51

Consider the groups $(\mathbb{Z}, +)$ of integers with addition and $(\mathbb{C} \setminus \{0\}, \cdot)$ of complex numbers with multiplication. From the lecture Grundlagen der Mathematik it is known that

$$\alpha : \mathbb{Z} \longrightarrow \mathbb{C} \setminus \{0\} : z \mapsto e^{\frac{i\cdot\pi}{2}\cdot z}$$

is a group homomorphism since the power law

$$e^{\frac{i\cdot\pi}{2}\cdot(z+z')} = e^{\frac{i\cdot\pi}{2}\cdot z} \cdot e^{\frac{i\cdot\pi}{2}\cdot z'}$$

holds true. A simple computation shows that

$$\operatorname{Im}(\alpha) = \{1, -1, i, -i\}$$

and

$$\operatorname{Ker}(\alpha) = 4 \cdot \mathbb{Z},$$

since $e^{\frac{i\cdot\pi}{2}\cdot z} = 1$ if and only if $\frac{z}{2}$ is a multiple of 2. The Homomorphism Theorem then implies

$$\mathbb{Z}_4 = \mathbb{Z}/4\mathbb{Z} = \mathbb{Z}/\operatorname{Ker}(\alpha) \cong \operatorname{Im}(\alpha) = \{1, -1, i, -i\},$$

where the group operation on the left hand side is the addition and on the right hand side it is multiplication

In order to understand the *well definedness* of the map $\widetilde{\alpha}$ in the Homomorphism Theorem in an example we recommend to note the following. In $\mathbb{Z}/4\mathbb{Z}$ the cosets $\overline{2}$ and $\overline{6}$ coincide. Therefore under the map

$$\widetilde{\alpha} : \mathbb{Z}/4\mathbb{Z} \longrightarrow \{1, -1, i, -i\} : \overline{z} \mapsto e^{\frac{i\cdot\pi}{2}\cdot z}$$

---

precisely this fact: $\overline{g} = \overline{h} \implies \alpha(g) = \alpha(h)$, or with the above notation

$$\overline{g} = \overline{h} \implies \widetilde{\alpha}(\overline{g}) = \widetilde{\alpha}(\overline{h}).$$

This looks close to the injectivity of a map, however, for the injectivity we had to check just the opposite implication!

also $\widetilde{\alpha}(\overline{2}) = \widetilde{\alpha(\overline{6})}$ has to hold true. In view of the definition of $\widetilde{\alpha}$ this means that necessarily $\alpha(2) = \alpha(6)$ must hold true. This, however, is correct since 2 and 6 differ by 4 and $e^{\frac{i\cdot\pi}{2}\cdot 4} = 1$.

Remark 4.52
Consider for $n \geq 2$ again the surjective group homomorphism (see Remark 3.20)

$$\mathrm{sgn} : \mathbb{S}_n \longrightarrow \{-1, 1\}.$$

The Homomorphism Theorem 4.50 implies in particular $|\mathbb{S}_n/\mathbb{A}_n| = |\{-1, 1\}| = 2$. Since by the Theorem of Lagrange 4.10 also $|\mathbb{S}_n/\mathbb{A}_n| = \frac{|\mathbb{S}_n|}{|\mathbb{A}_n|}$ holds true we get with Proposition 3.11 the following equation:

$$|\mathbb{A}_n| = \frac{n!}{2}.$$

The following isomorphism theorems are easy applications of the above Homomorphism Theorem.

Theorem 4.53 (1st Isomorphism Theorem)
If G is a group, $U \leq G$ and $N \trianglelefteq G$. Then

$$U/U \cap N \cong UN/N.$$

Proof: We leave the proof to the reader as an exercise. $\square$

Theorem 4.54 (2nd Isomorphism Theorem)
Let G be a group, $M \subseteq N \subseteq G$ be two normal subgroups of G. Then also N/M is a normal subgroup of G/M and we have

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

Proof: We consider the following map

$$\beta : G/M \to G/N : gM \mapsto gN,$$

and show that it is an epimorphism with kernel N/M. This in particular shows that N/M is a normal subgroup of G/M.[23]

***Step 0*:** β is well-defined.

Since we define the map β via the choice of a (non-unique) representative of the coset, we have to show that β is well-defined, i. e. that the definition is independent of the chosen representative. Let therefore $gM = g'M$, then $g^{-1} \cdot g' \in M \subseteq N$, and thus $= gN = g'N$, i. e. gN does not depend on the representative of gM.

***Step 1*:** β is a homomorphism.

Let $gM, g'M \in G/M$ be given. Then

$$\beta\big(gM \cdot g'M\big) = \beta\big(gg'M\big) = gg'N = gN \cdot g'N = \beta(gM) \cdot \beta\big(g'M\big).$$

[23]Note that M is obviously a normal subgroup of N and hence the quotient N/M is indeed defined and it coincides with the set of left cosets of M in G of the form nM with $n \in N$.

***Step 2*:** β is surjective.

Let $gN \in G/N$ be given. Then $gN = \beta(gM) \in \mathrm{Im}(\beta)$, so that β is surjective.

***Step 3*:** $\mathrm{Ker}(\beta) = N/M$.

$gM \in \mathrm{Ker}(\beta)$ if and only if $gN = N$ if and only if $g \in N$ if and only if $gM \in N/M$. $\square$

Exercise 4.55
Prove Theorem 4.53

With the help of the Homomorphism Theorem and the Theorem of Lagrange one can solve the following exercise.

Exercise 4.56
Let $(G, \cdot)$ and $(H, *)$ be two finite groups of coprime order. Show that there is precisely one group homomorphism $\alpha : G \longrightarrow H$.

Exercise 4.57
Let $(G, \cdot)$ be a group.

a. If $g, h \in G$ with $o(g) = o(h) = p$, where p a prime number, then either $\langle g \rangle = \langle h \rangle$ or $\langle g \rangle \cap \langle h \rangle = \{e\}$.
b. If $|G| = 10$ the there are two elements $g, h \in G$ with:
   - $o(g) = 2$,
   - $o(h) = 5$,
   - $\langle h \rangle \trianglelefteq G$,
   - $\langle g \rangle \cdot \langle h \rangle = G$.

Hint, for part b. show first that neither of the following two possibilities can occur: 1st $o(k) = 2$ for all $e \neq k \in G$, 2nd $o(k) = 5$ for all $e \neq k \in G$.

A group G as in Exercise 4.57 b. is called a semidirect product of $\langle g \rangle$ and $\langle h \rangle$. One can show that if $g \cdot h = h \cdot g$ then G is isomorphic to $\mathbb{Z}_{10}$ while otherwise G is isomorphic to $\mathbb{D}_{10}$.

Exercise 4.58
Find all group homomorphisms $\alpha : \mathbb{Z}_{10} \longrightarrow \mathbb{Z}_n$ with $n \in \{6, 13\}$.

E) Cyclic Groups

We want to close this section with the classification of all cyclic groups.

Theorem 4.59
Let $G = \langle g \rangle$ be a cyclic group.

a. *If $|G| = \infty$ then we have the group isomorphism*

$$\alpha : \mathbb{Z} \xrightarrow{\cong} G : z \mapsto g^z.$$

b. *If* $|G| = n < \infty$ *then we have the group isomorphism*

$$\overline{\alpha} : \mathbb{Z}_n \xrightarrow{\cong} G : \overline{z} \mapsto g^z.$$

Proof: For the map

$$\alpha : \mathbb{Z} \xrightarrow{\cong} G : z \mapsto g^z$$

and two integers $x, y \in \mathbb{Z}$ we have

$$\alpha(a + y) = g^{x+y} = g^x \cdot g^y = \alpha(x) \cdot \alpha(y).$$

Hence, α is a group homomorphism and

$$\mathrm{Im}(\alpha) = \{g^z \mid z \in \mathbb{Z}\} = \langle g \rangle = G,$$

i.e. α is surjective.

If $|G| = o(g) = \infty$ then

$$\{0\} = \{z \in \mathbb{Z} \mid g^z = e\} = \mathrm{Ker}(\alpha)$$

by Remark 4.12, i.e. in this case α is also injective.

If $|G| = o(g) = n < \infty$ then by Exercise 1.55

$$\mathrm{Ker}(\alpha) = \{z \in \mathbb{Z} \mid g^z = e\} = n\mathbb{Z}.$$

The Homomorphism Theorem therefore implies that the map $\overline{\alpha}$ is a group isomorphism. □

The classification of cyclic groups can be applied to compute the order of g^n from the order of g.

Corollary 4.60
If $(G, \cdot)$ *is a group,* $g \in G$ *with* $o(g) < \infty$ *and* $0 \neq m \in \mathbb{Z}$. *Then*

$$o\big(g^m\big) = \frac{\mathrm{lcm}\,\big(m, o(g)\big)}{|m|}.$$

Proof: Let $n = o(g)$ then

$$\overline{\alpha} : \mathbb{Z}_n \longrightarrow \langle g \rangle : \overline{z} \mapsto g^z$$

is a group isomorphism by Theorem 4.59. Exercise 1.61 implies therefore that

$$o\big(g^m\big) = o\Big(\overline{\alpha}(\overline{m})\Big) = o\big(\overline{m}\big).$$

If $m > 0$ then the claim follows from Corollary 4.44. If $m < 0$ then $-m > 0$ and we get analogously

$$o\big(g^{-m}\big) = \frac{\mathrm{lcm}(-m, n)}{-m} = \frac{\mathrm{lcm}(m, n)}{|m|}.$$

Since moreover the order of an element and its inverse coincide the claim follows also in the case $m < 0$. □

Corollary 4.61
If $G = \langle g \rangle$ *is a cyclic group of order* $|G| = n < \infty$ *then:*

$$U \leq G \iff \exists\, m \in \{1, \ldots, n\} \text{ with } m \text{ divides } n \;:\; U = \langle g^m \rangle.$$

For such a subgroup the following holds true:

$$|\langle g^m \rangle| = \frac{n}{m}.$$

In particular, G *has precisely one subgroup of order* d *for each divisor* d *of* n*.*

Proof: By Theorem 4.59 the map

$$\overline{\alpha} : \mathbb{Z}_n \longrightarrow G : \overline{z} \mapsto g^z$$

is a group isomorphism, so that the first assertion follows from Corollary 4.42. The claim on the order the follows from Corollary 4.60, since $\mathrm{lcm}(m, n) = n$. Finally we should note that with m also $\frac{n}{m}$ runs through all divisors of n. □

Since the subgroups of $\mathbb{Z}$ are cyclic by Proposition 1.39 we get the following statement by Theorem 4.59 and Corollary 4.42.

Corollary 4.62
Each subgroup of a cyclic group is cyclic.

Exercise 4.63
If $(G, \cdot)$ is a group and $p = |G|$ a prime number then G is isomorphic to $(\mathbb{Z}_p, +)$.

Exercise 4.64
Let $(G, \cdot)$ be a group, $g \in G$ and $n \in \mathbb{Z}_{>0}$. Show there is a group homomorphism $\alpha : \mathbb{Z}_n \longrightarrow G$ with $\alpha(\overline{1}) = g$ if and only if the order of g is a divisor of n.

Exercise 4.65
Find all automorphisms of the group $(\mathbb{Z}_{10}, +)$.

5 Check Digit Codes

Nowadays products in shops all carry bar codes and are identified by them. Moreover, at the cash desk the bar code is scanned or typed in and that way you get charged the price. Sometimes the bar codes are not recognised correctly or the wrong number has been typed in. However, the error is recognised by the machine and the bar code is not accepted.

A) Have you ever wondered how it comes, that you are always charged the right price?

Well, the machine looks the bar code up in some data base, and if the incorrect bar code was contained in that data base as well, then the machine could not possibly detect any error. So, when assigning bar codes, you have to make sure that no bar codes which - in a certain sense - are too similar are in the data base.

Is this difficult? Well, to decide on that question we should know, what bar codes in principle look like!

Bar codes are also called **EAN-13** codes, where EAN is short for European Article Number, and they consist of a thirteen digit number. The first 2 to 3 digits stand for the organisation which assigned the numbers to the producer, some of the next digits identify this producer and so on. So, the digits are not really arbitrary digits. In particular, for a fixed producer a large part of the bar code will always be the same. I. e. the numbers will have to be similar!

How can we get along with that problem?

Idea: Store some *redundant information* which is not needed to identify the article, but only to detect possible errors.

In the case of the EAN-13 only 12 digits characterise the article. Digit no. 13 is a so called **check digit**.

B) How is the check digit related to the (real) article number?

Basic Idea: It should be possible to calculate the check digit from the remaining digits in an easy way, but such that (common) errors are possibly detected.

First Idea: Repeat the whole number! This is a bit too much redundancy and increases the risk of falsely scanned numbers.

Second Idea: Take the cross sum of the digits of the real product number as check “digit”.

E. g. if the product number is 013412547180, then the check digit would be

$$0+1+3+4+1+2+5+4+7+1+8+0=36.$$

This will usually be several digits long, and is still too much redundancy.

Third Idea: Let's just take the last digit of the cross sum!

E. g. in the above example the check digit would then be 6.

This can be formulated in a more mathematical way by saying that

we take the remainder of the cross sum by division with remainder modulo 10.

And that's where groups come into play as a nice way to formulate the procedure. We may identify the digits $0, \ldots, 9$ with the elements of the additive group $(\mathbb{Z}/10\mathbb{Z}, +)$, just via the map

$$\{0, \ldots, 9\} \to \mathbb{Z}/10\mathbb{Z} : a \mapsto \overline{a} = a + 10\mathbb{Z} = \{a + 10z \mid z \in \mathbb{Z}\},$$

i. e. identifying the digit with the residue class represented by the number. Viewing the digits in the article number as elements of $\mathbb{Z}/10\mathbb{Z}$ that way, the check digit becomes just the sum of the "digits".

E. g. $\overline{0} + \overline{1} + \overline{3} + \overline{4} + \overline{1} + \overline{2} + \overline{5} + \overline{4} + \overline{7} + \overline{1} + \overline{8} + \overline{0} = \overline{36} = \overline{6}$.

C) Does this allow to detect errors? Otherwise it is of no use.

Certainly we will not be able to detect all errors, thus we have to distinguish certain types of errors! Some statistics tell us that the following two types are the most common ones.

Type I: "Single Digit Errors" – i. e. just one digit is wrong. These are roughly 80% of the occuring errors.

Type II: "Neighbour Transpositions" – i. e. two neighbouring digits have been interchanged. These are about 10% of the errors.

It is fairly obvious that the cross-sum-mod-10-approach cannot detect errors of Type II, since the addition in $\mathbb{Z}/10\mathbb{Z}$ is commutative. However, does it detect errors of Type I?

Suppose the correct number was $a_1a_2 \cdots a_{13}$ and instead of some a_i we read $a_i' \in \{0, \ldots, 9\}$ with $a_i \neq a_i'$. Then

$$\overline{a_{13}} - \left(\sum_{j \neq i,13} \overline{a_j} + \overline{a_i'} \right) = \sum_{j=1}^{12} \overline{a_j} - \left(\sum_{j \neq i,13} \overline{a_j} + \overline{a_i'} \right) = \overline{a_i - a_i'} \neq \overline{0}, \quad (23)$$

since $a_i - a_i'$ is number between -9 and 9 which is non-zero and thus 10 does not divide $a_i - a_i'$. That means "Single Digit Errors" are detected.

D) Back to EAN-13.

The encoding of EAN-13 is, however, slightly different. The check digit in $a_1a_2 \cdots a_{13}$ satisfies

$$\overline{a_{13}} = (-1) \cdot \overline{a_1} + (-3) \cdot \overline{a_2} + (-1) \cdot \overline{a_3} + (-3) \cdot \overline{a_4} + \ldots + (-1) \cdot \overline{a_{13}}$$

or equivalently

$$\overline{a_1} + 3 \cdot \overline{a_2} + \overline{a_3} + \ldots + \overline{a_{13}} = \overline{0}.$$

We call these equations **check digit equations**.

Does this still detect errors of Type I?

Let's go back to Equation (23) for this. The question finally comes down to checking whether $a_i \neq a_i'$ implies that $\overline{a_i - a_i'}$ and $\overline{3 \cdot (a_i - a_i')}$ are not equal to $\overline{0}$, which is the case since $a_i - a_i'$ is not divisible by 10 and thus also three times this number is not. Thus we are lucky.

How about errors of Type II?

If a_i and a_{i+1} have been interchanged, then this comes down to the question whether

$$\begin{aligned} & 3 \cdot \overline{a_i} + \overline{a_{i+1}} = 3 \cdot \overline{a_{i+1}} + \overline{a_i} \\ \Leftrightarrow \quad & \overline{2 \cdot (a_i - a_{i+1})} = \overline{0} \\ \Leftrightarrow \quad & 5 \mid a_i - a_{i+1}. \end{aligned}$$

Thus even errors of Type II will quite frequently be detected, but not all of them. We achieved this by multiplying the digits in the cross sum by certain weights w_i – here $w_i = 1$ and $w_i = 3$.

E) WHICH WEIGHTS w_i WOULD HAVE BEEN SUITABLE IN THE CHECK DIGIT EQUATION IN ORDER NOT TO LOOSE THE PROPERTY THAT ERRORS OF TYPE I ARE DETECTED?

The important point was that

$$\overline{a_i} \neq \overline{a_i'} \quad \Rightarrow \quad \omega_i \cdot \overline{a_i} \neq \omega_i \cdot \overline{a_i'},$$

i. e. that the map

$$\mu_{\omega_i} : \mathbb{Z}/10\mathbb{Z} \to \mathbb{Z}/10\mathbb{Z} : \overline{a} \mapsto \omega_i \cdot \overline{a}$$

is injective, and hence bijective since $\mathbb{Z}/10\mathbb{Z}$ is a finite set. In other words, μ_{ω_i} is a permutation of the set $\mathbb{Z}/10\mathbb{Z}$.

This leads to the following generalisation and definition.

Definition 5.1
Let $(G, \cdot)$ be a group, $g_0 \in G$ a fixed element, and let $\pi_1, \ldots, \pi_n \in \mathrm{Sym}(G)$ be permutations.

a. We call

$$C = C_G(\pi_1, \ldots, \pi_n, g_0) = \left\{ (g_1, \ldots, g_n)^t \in G^n \;\middle|\; \pi_1(g_1) \cdots \pi_n(g_n) = g_0 \right\}$$

a *check digit code* (CDC) of *length* n on the *alphabet* G.

b. We say that C detects errors of Type I if and only if $(g_1, \ldots, g_n)^t \in C$ and $g_i' \in G$ with $g_i' \neq g_i$ implies that $(g_1, \ldots, g_{i-1}, g_i', g_{i+1}, \ldots, g_n)^t \notin C$.

c. We say that C detects errors of Type II if and only if $(g_1, \ldots, g_n)^t \in C$ with $g_i \neq g_{i+1}$ implies that $(g_1, \ldots, g_{i-1}, g_{i+1}, g_i, g_{i+2}, \ldots, g_n)^t \notin C$.

Example 5.2 (EAN-13)
Let $(G,\cdot) = (\mathbb{Z}/10\mathbb{Z}, +)$, $g_0 = \overline{0}$, $n = 13$, $\pi_i = \mu_1$ if i is odd and $\pi_i = \mu_3$ if i is even. This then describes the EAN-13 code $C = C_{\mathbb{Z}/10\mathbb{Z}}(\mu_1, \mu_3, \ldots, \mu_1, \overline{0})$.

Actually, $C = \ker(\phi)$, where $\phi : (\mathbb{Z}/10\mathbb{Z})^{13} \longrightarrow \mathbb{Z}/10\mathbb{Z}$ is the group homomorphism defined by multiplication with the matrix $\left(\overline{1}, \overline{3}, \overline{1}, \ldots, \overline{1}\right)$.

Having introduced check digit codes over arbitrary groups it would be nice to know something about their error detecting properties.

Proposition 5.3 (Error Detecting Properties)
Let $C = C_G(\pi_1, \ldots, \pi_n, g_0)$ be a CDC over the alphabet $(G,\cdot)$.

a. C *detects errors of Type I.*
b. *If $n \geq 3$, then C detects errors of Type II if and only if $\forall\ i = 1, \ldots, n-1$, $\forall\ g, h \in G$ s. t. $g \neq h$:*

$$g \cdot \left(\pi_{i+1} \circ \pi_i^{-1}\right)(h) \neq h \cdot \left(\pi_{i+1} \circ \pi_i^{-1}\right)(g).$$

Proof: a. Let $(g_1, \ldots, g_n)^t \in C$, $g_i' \in G$ such that $g_i' \neq g_i$, and suppose $(g_1, \ldots, g_i', \ldots, g_n)^t \in C$. Then

$$\pi_1(g_1) \cdots \pi_n(g_n) = g_0 = \pi_1(g_1) \cdots \pi_i\left(g_i'\right) \cdots \pi_n(g_n).$$

By the cancellation law we thus deduce that

$$\pi_i(g_i) = \pi_i\left(g_i'\right).$$

But then also $g_i = g_i'$, since π_i is injective. This, however, is a contradiction to our assumption.

b. Let's first assume that the condition of the proposition is satisfied and let's show that then C detects errors of Type II. For this let $(g_1, \ldots, g_n)^t \in C$ be given with $g_i \neq g_{i+1}$ and set $g = \pi_i(g_i)$ and $h = \pi_i(g_{i+1})$. Since π_i is injective we have $g \neq h$. Thus by the condition of the proposition we also have

$$\pi_i(g_i) \cdot \pi_{i+1}(g_{i+1}) = g \cdot \left(\pi_{i+1} \circ \pi_i^{-1}\right)(h) \neq h \cdot \left(\pi_{i+1} \circ \pi_i^{-1}\right)(g) = \pi_i(g_{i+1}) \cdot \pi_{i+1}(g_i).$$

Multiplying both sides with the same element of G the inequality is preserved and we get

$$\pi_1(g_1) \cdots \pi_i(g_i) \cdot \pi_{i+1}(g_{i+1}) \cdots \pi_n(g_n) \neq \pi_1(g_1) \cdots \pi_i(g_{i+1}) \cdot \pi_{i+1}(g_i) \cdots \pi_n(g_n).$$

This means that C detects errors of Type II.

Let's now suppose that C detects errors of Type II and then prove the above condition. For this let $g, h \in G$ with $g \neq h$, and set $g_i = \pi_i^{-1}(g)$ and $g_{i+1} = \pi_i^{-1}(h)$. Since π_i is bijective $g_i \neq g_{i+1}$. Choose now $g_j \in G$, $j \neq i, i+1$ such that $(g_1, \ldots, g_n)^t \in C$ (here we need $n \geq 3$). Thus by assumption

$$(g_1, \ldots, g_{i+1}, g_i, \ldots, g_n)^t \notin C.$$

But then

$$\pi_1(g_1)\cdots\pi_n(g_n) = g_0 \neq \pi_1(g_1)\cdots\pi_i(g_{i+1})\cdot\pi_{i+1}(g_i)\cdots\pi_n(g_n).$$

Using the cancellation law we derive

$$g\cdot\big(\pi_{i+1}\circ\pi_i^{-1}\big)(h) = \pi_i(g_i)\cdot\pi_{i+1}(g_{i+1}) \neq \pi_i(g_{i+1})\cdot\pi_{i+1}(g_i) = h\cdot\big(\pi_{i+1}\circ\pi_i^{-1}\big)(g).$$

This finishes the proof.

□

Note: If $(G,\cdot)$ is *abelian* and $\mathrm{inv} : G \longrightarrow G : g \mapsto g^{-1}$ denotes the inversion map, then the condition in Proposition 5.3 comes down to

$$g\cdot\big(\mathrm{inv}\circ\pi_{i+1}\circ\pi_i^{-1}\big)(g) \neq h\cdot\big(\mathrm{inv}\circ\pi_{i+1}\circ\pi_i^{-1}\big)(h). \tag{24}$$

Since $\mathrm{inv}\circ\pi_{i+1}\circ\pi_i^{-1} \in \mathrm{Sym}(G)$ is a permutation of G, it seems that maps of the form $g \mapsto g\pi(g)$ for some permutation $\pi \in \mathrm{Sym}(G)$ are connected to the error detecting properties of codes.

Definition 5.4
Let $(G,\cdot)$ be a group and $\pi \in \mathrm{Sym}(G)$. We call π a *complete mapping* if and only if the map

$$\pi^* : G \longrightarrow G : g \mapsto g\cdot\pi(g)$$

is again a permutation of G.

So far we know how to check whether a given CDC detects errors of Typer II or not, but we have no means to find such a code – or possibly to decide that their is non.

Corollary 5.5
Let $(G,\cdot)$ be a finite abelian group, $n \geq 3$. Then there is a CDC of length n which detects errors of Type II if and only if G admits a complete mapping.

Proof: Let's first suppose that G admits a complete mapping $\pi \in \mathrm{Sym}(G)$. Set $g_0 = e_G$ and $\pi_i = (\mathrm{inv}\circ\pi)^i$ for $i = 1,\ldots,n$.

Claim: $C = C_G(\pi_1,\ldots,\pi_n,g_0)$ detects errors of Type II.

For this we only have to check that Equation (24) is satisfied. Let $g,h \in G$ such that $g \neq h$. Then

$$\begin{aligned} g\cdot\big(\mathrm{inv}\circ\pi_{i+1}\circ\pi_i^{-1}\big)(g) &= g\cdot\big(\mathrm{inv}\circ(\mathrm{inv}\circ\pi)^{i+1-i}\big)(g) = g\cdot\pi(g) = \pi^*(g)\\ \neq \pi^*(h) = h\cdot\pi(h) &= h\cdot\big(\mathrm{inv}\circ(\mathrm{inv}\circ\pi)^{i+1-i}\big)(h) = h\cdot\big(\mathrm{inv}\circ\pi_{i+1}\circ\pi_i^{-1}\big)(h). \end{aligned}$$

Thus Equation (24) is fulfilled.

Let's now suppose that there is a CDC $C_G(\pi_1,\ldots,\pi_n,g_0)$ which detects errors of Type II. We define $\pi = \mathrm{inv}\circ\pi_2\circ\pi_1^{-1} \in \mathrm{Sym}(G)$ and we claim that this is then a complete mapping. In order to check this we let $g,h \in G$ such that $g \neq h$. Thus by Equation (24) we have

$$\pi^*(g) = g\cdot\pi(g) = g\cdot\big(\mathrm{inv}\circ\pi_2\circ\pi_1^{-1}\big)(g) \neq h\cdot\big(\mathrm{inv}\circ\pi_2\circ\pi_1^{-1}\big)(h) = h\cdot\pi(h) = \pi^*(h).$$

Hence π^* is injective and thus bijective, since G is finite. But then π is a complete mapping. $\square$

Remark

a. If $|G| = 2 \cdot m$ with m odd, then there exists *no* complete mapping on G.[24]

In particular, there is no CDC on $\mathbb{Z}/10\mathbb{Z}$ which detects all errors of Type II.

b. If $|G|$ is odd, then the identity mapping id_G is a complete mapping.

Proof: Let $|G| = 2m+1$, then by the Theorem of Lagrange we have $e_G = g^{|G|} = g^{2m+1}$. Multiplying by g we get $\left(g^{m+1}\right)^2 = g$, and thus the mapping $\mathrm{id}_G^* : g \mapsto g \cdot \mathrm{id}_G(g) = g^2$ is surjective. But since G is finite, it is then bijective. $\square$

c. **Problem:** There is no CDC on $(\mathbb{Z}/10\mathbb{Z}, +)$ which detects errors of Type II! How can we deal with that?

Solution 1: Use an odd number of digits, i. e. calculate over $\mathbb{Z}/m\mathbb{Z}$ with an odd m.

E. g. the ISBN code works over $(\mathbb{Z}/11\mathbb{Z}, +)$, where the element $\overline{10} = 10 + 11\mathbb{Z}$ is denoted by X and is only used as check digit. The ISBN code is a $C_{\mathbb{Z}/11\mathbb{Z}}(\pi_1, \ldots, \pi_{10}, \overline{0})$ code, where $\pi_i : \mathbb{Z}/11\mathbb{Z} \to \mathbb{Z}/11\mathbb{Z} : \overline{a} \mapsto i \cdot \overline{a}$. We leave it as an exercise to check that the code actually detects errors of Type II. You only have to check that Equation (24) is satisfied.

Solution 2: Use a non-abelian group with ten elements! There the non-existence of a complete mapping is not related to the error detecting property.

Example 5.7 (German Currency)
The check digits of the serial numbers of the German currency where actually encoded by a $C_{\mathbb{D}_{10}}\big(\pi_1, \ldots, \pi_{10}, \mathrm{id}_{\mathbb{D}_{10}}, (1)\big)$ code.

Consider the *dihedral group*

$$\mathbb{D}_{10} = \big\langle (1\ 2\ 3\ 4\ 5), (1\ 5)(2\ 4) \big\rangle \leq \mathbb{S}_5 = \mathrm{Sym}\left(\{1, \ldots, 5\}\right).$$

In the exercises you show that, setting $\sigma = (1\ 2\ 3\ 4\ 5)$ and $\tau = (1\ 5)(2\ 4)$, we may describe $\mathbb{D}_{10}$ as the set

$$\mathbb{D}_{10} = \left\{\sigma^0 = (1), \sigma^1, \ldots, \sigma^4, \tau \circ \sigma^0 = \tau, \tau \circ \sigma^1, \ldots, \tau \circ \sigma^4\right\}.$$

And since $\tau \circ \sigma = \sigma^{-1} \circ \tau \neq \sigma \circ \tau$, the group is indeed not abelian.

Verhoeff showed in [Ver75] that the permutation $\pi : \mathbb{D}_{10} \to \mathbb{D}_{10}$ of $\mathbb{D}_{10}$ defined by

x	σ^0	σ^1	σ^2	σ^3	σ^4	$\tau \circ \sigma^0$	$\tau \circ \sigma^1$	$\tau \circ \sigma^2$	$\tau \circ \sigma^3$	$\tau \circ \sigma^4$
$\pi(x)$	σ^1	$\tau \circ \sigma^0$	$\tau \circ \sigma^2$	$\tau \circ \sigma$	σ^2	$\tau \circ \sigma^3$	σ^3	σ^0	$\tau \circ \sigma^4$	σ^4

satisfies that $g, h \in \mathbb{D}_{10}$ with $g \neq h$ implies $g \circ \pi(h) \neq h \circ \pi(g)$. Hence, setting $\pi_i = \pi^i \in \mathrm{Sym}(\mathbb{D}_{10})$, the code $C_{\mathbb{D}_{10}}\big(\pi_1, \ldots, \pi_{10}, (1)\big)$ detects errors of Type II by Proposition 5.3.

[24]The proof is elementary, but lengthy. We refer the reader to [Sie81].

Of course for the serial numbers on the German currency they did not use such fancy symbols like σ. They used the usual 10 digits and in addition 10 letters. However, they were identified with the elements in $\mathbb{D}_{10}$ in the following way

σ^0	σ^1	σ^2	σ^3	σ^4	$\tau\circ\sigma^0$	$\tau\circ\sigma^1$	$\tau\circ\sigma^2$	$\tau\circ\sigma^3$	$\tau\circ\sigma^4$
0	1	2	3	4	5	6	7	8	9
A	D	G	K	L	N	S	U	Y	Z.

Thus, if you wanted to check whether a serial number on a German bank note was valid, you replaced the digits and letters by the appropriate elements of $\mathbb{D}_{10}$ and looked whether this element belonged to $C_{\mathbb{D}_{10}}\big(\pi_1,\ldots,\pi_{10},\mathrm{id}_{\mathbb{D}_{10}},(1)\big)$.

Exercise 5.8

Check if $AA6186305Z2$ is a valid serial number for a German bank note.

Question: Could we have used some other group with 10 elements as alphabet?

Answer: No! Not really. The group only matters up to isomorphism, and one can show that up to isomorphism there are only two groups with 10 elements – $(\mathbb{Z}/10\mathbb{Z},+)$ and $(\mathbb{D}_{10},\circ)$.

6 Rings and Fields

A) Rings and Fields

In Section 1 we introduced the mathematical structure of a *group*, and our first examples were the additive groups $(\mathbb{Z}, +)$ and $(\mathbb{Q}, +)$ of integers and rational numbers respectively. On these two sets we have besides the operation of addition a second binary operation, the multiplication, and in both cases there are again interesting rules satisfied. E.g. the set $(\mathbb{Q} \setminus \{0\}, \cdot)$ is again a group while the set $(\mathbb{Z} \setminus \{0\}, \cdot)$ comes short of this property only by missing the multiplicative inverses. We now want to generalise these examples which leads to the following definition.

Definition 6.1

a. A *ring with one* is a triple $(R, +, \cdot)$ consisting of a set R together with two binary operations

$$+ : R \times R \to R : (a, b) \mapsto a + b, \qquad (\text{“}\textit{addition}\text{”})$$

and

$$\cdot : R \times R \to R : (a, b) \mapsto a \cdot b, \qquad (\text{“}\textit{multiplication}\text{”})$$

such that the following axioms are fulfilled:

(i) $(R, +)$ is an abelian group (with neutral element 0_R).

(ii) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in R$. (“*associativity of the multiplication*”)

(iii) there is an element $1_R \in R$ with $1_R \cdot a = a \cdot 1_R = a$ for all $a \in R$. (“*the 1-element*”)

(iv) $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$ for all $a, b, c \in R$. (“*distributivity*”)

b. A ring with one $(R, +, \cdot)$ is called *commutative* if $a \cdot b = b \cdot a$ for all $a, b \in R$.

c. If $(R, +, \cdot)$ is a ring with one then $a \in R$ is a *unit* or *invertible* in R if there is an $a' \in R$ with $a \cdot a' = a' \cdot a = 1_R$. We denote by

$$R^* = \{a \in R \mid a \text{ is unit}\}$$

the set units in R.

d. A commutative ring with one $(R, +, \cdot)$ is called a *field* if $1_R \in R^* = R \setminus \{0\}$.

Remark 6.2

We will always denote the addition in rings by the symbol $+$ and the multiplication by the symbol $\cdot$, even if we consider several rings at the same time. We will therefore usually talk about a ring R rather than the ring $(R, +, \cdot)$, omitting the ring operations. Moreover, we will in general write ab when we mean $a \cdot b$.

The neutral element of $(R, +)$ will always be denoted by 0_R or simply by 0, and we call it the 0-*element* of R; similarly the 1-element will be denoted by 1_R or simply by 1.

If R is a ring and $a, b \in R$ we abbreviate $a + (-b)$ to $a - b$.

The 1-element in R is uniquely determined since if $1_R, 1'_R \in R$ are two elements with the property of the 1-element then $1_R = 1_R \cdot 1'_R = 1'_R$.

If $a \in R$ is a unit and $a', a'' \in R$ with $a \cdot a' = a' \cdot a = 1_R$ and $a \cdot a'' = a'' \cdot a = 1_R$ then

$$a' = 1_R \cdot a' = (a'' \cdot a) \cdot a' = a'' \cdot (a \cdot a') = a'' \cdot 1_R = a''.$$

The inverse a' to a is therefore uniquely determined and it is denoted by a^{-1} or $\frac{1}{a}$.

Note that the definition implies right away that $(R^*, \cdot)$ is a group, the so called *group of units* of the ring R.

Taking the rules in Lemma 6.5 into account we can show easily that a triple $(K, +, \cdot)$ is a *field* if and only if

a. $(K, +)$ is an abelian group.
b. $(K \setminus \{0\}, \cdot)$ is an abelian group.
c. For all $a, b, c \in K$ we have $a \cdot (b + c) = a \cdot b + a \cdot c$.

□

Example 6.3

a. $(\mathbb{Z}, +, \cdot)$ with the usual addition and multiplication is a commutative ring with one, but it is not a field.
b. $(\mathbb{Q}, +, \cdot)$ and $(\mathbb{R}, +, \cdot)$ with the usual addition and multiplication are fields.
c. In the Lecture Grundlagen der Mathematik the set $\mathbb{C} = \{(x, y) \mid x, y \in \mathbb{R}\}$ was considered together with the two binary operations

$$+ : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C} : \big((x, y), (x', y')\big) \mapsto (x + x', y + y')$$

and

$$\cdot : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C} : \big((x, y), (x', y')\big) \mapsto (x \cdot x' - y \cdot y', x \cdot y' + x' \cdot y).$$

Moreover, it was shown that $(\mathbb{C}, +, \cdot)$ is a field – the *field of complex numbers* For the elements of $\mathbb{C}$ the following notation is common: $(x, y) = x + iy$ with $i^2 = -1$. We will suppose for the remaining part of the lecture that the complex numbers and their properties are known.

d. If M is an arbitrary set and $(R, +, \cdot)$ is a ring with one then so is

$$R^M := \{f \mid f : M \to R \text{ is a map}\}$$

with the pointwise operations

$$+ : R^M \times R^M \to R^M : (f, g) \mapsto \big(f + g : M \to R : x \mapsto f(x) + g(x)\big),$$

and

$$\cdot : R^M \times R^M \to R^M : (f, g) \mapsto \big(f \cdot g : M \to R : x \mapsto f(x) \cdot g(x)\big).$$

The constant function $0 : M \to R : x \mapsto 0_R$ is the neutral element of the addition and the constant function $1 : M \to R : x \mapsto 1_R$ is the 1-element as one can show with a bit of effort.

e. In Exercise 1.21 we introduced the set

$$\mathrm{Mat}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; a, b, c, d \in \mathbb{R} \right\}$$

of real 2×2-matrices and for two matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in \mathrm{Mat}_2(\mathbb{R})$$

we defined their product as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a \cdot a' + b \cdot c' & a \cdot b' + b \cdot d' \\ c \cdot a' + d \cdot c' & c \cdot b' + d \cdot d' \end{pmatrix}.$$

If we define moreover

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a + a' & b + b' \\ c + c' & d + d' \end{pmatrix},$$

then only a bit of patience is necessary to show that $(\mathrm{Mat}_2(\mathbb{R}), +, \cdot)$ is a ring with 1-element

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This ring is not commutative since

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

In the lecture Grundlagen der Mathematik this example will be generalised to matrices of arbitrary size $n \times n$ for $n \geq 1$ whose entries belong to an arbitrary field K, and the proof that we get this way a commutative ring with one will be proved in a much more elegant way.

Example 6.4

Let $(R, +, \cdot)$ and $(S, +, \cdot)$ be two commutative rings with one. The Cartesian product $R \times S$ is a commutative ring with one by the componentwise operations

$$(r, s) + (r', s') := (r + r', s + s')$$

and

$$(r, s) \cdot (r', s') := (r \cdot r', s \cdot s'),$$

and the 1-element of $R \times S$ is $(1_R, 1_S)$. Checking the axioms is an easy application of the definition.

For the units in $R \times S$ we have

$$(R \times S)^* = R^* \times S^*,$$

since

$$(1_R, 1_S) = (r, s) \cdot (r', s') = (r \cdot r', s \cdot s') \iff 1_R = r \cdot r' \text{ and } 1_S = s \cdot s'.$$

In the same way the Cartesian product of an arbitrary number of commutative rings with one is again a commutative ring with one by componentwise operations, and the group of units of the new rings is just the Cartesian product of the groups of units of the individual rings.

We now want to give some rules for computing in rings.

Lemma 6.5 (Computational Rules)
Let R *be a ring with one. For* $a, b, c \in R$ *we have*

a. $-(-a) = a$.

b. $a + b = c \iff a = c - b$.

c. $-(a + b) = -a - b$.

d. $0 \cdot a = a \cdot 0 = 0$.

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

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

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

h. *For* $a \in R^*$ *we have* $a^{-1} \in R^*$ *and* $\left(a^{-1}\right)^{-1} = a$.

i. *If* $1_R = 0_R$ *then* $R = \{0_R\}$ *the* nullring.

Proof: The statements a., b. and c. follow right away from Lemma 1.6.

d. For $a \in R$ we have $0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$. Hence it follows that $0 \cdot a = 0$ by applying the cancellation rules in $(R, +)$. Analogously we check that $a \cdot 0 = 0$.

e. For $a, b \in R$ we have by d.:

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

Hence, $-(a \cdot b) = (-a) \cdot b$. The equality of these two terms to $a \cdot (-b)$ is shown in the same way.

f. For $a, b \in R$ we get by a. and e.:

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

g. For $a, b, c \in R$ Part e. implies:

$$a \cdot (b - c) = a \cdot b + a \cdot (-c) = a \cdot b + \big(-(a \cdot c)\big) = a \cdot b - a \cdot c.$$

h. If $a \in R^*$ is a unit with inverse a^{-1}. Then by definition a is an inverse of a^{-1}. In particular a^{-1} is a unit. The uniqueness of the inverse (see Remark 6.2) shows then that $a = (a^{-1})^{-1}$.

i. If $a \in R$ then $a = 1_R \cdot a = 0_R \cdot a = 0_R$.

□

An important class of commutative rings with one are the so called formal power series rings.

Definition 6.6
Let R be a commutative ring with one. If $a_k \in R$ for $k \in \mathbb{N}$ then the map

$$\mathbb{N} \longrightarrow R : k \mapsto a_k$$

can be represented by the expression $\sum_{k=0}^{\infty} a_k \cdot t^k$. The set

$$R[[t]] := \left\{ \sum_{k=0}^{\infty} b_k \cdot t^k \;\middle|\; b_k \in R \right\}$$

is the set of all such maps from $\mathbb{N}$ to R. We call the elements of $R[[t]]$ *formal power series*, and $R[[t]]$ is called the *ring of formal power series* over R in the indeterminate t.

For two formal power series $\sum_{i=0}^{\infty} a_i \cdot t^i, \sum_{j=0}^{\infty} b_j \cdot t^j \in R[[t]]$ we define

$$\sum_{i=0}^{\infty} a_i \cdot t^i \;+\; \sum_{i=0}^{\infty} b_i \cdot t^i \;:=\; \sum_{i=0}^{\infty} (a_i + b_i) \cdot t^i \in R[[t]]$$

and

$$\sum_{i=0}^{\infty} a_i \cdot t^i \;\cdot\; \sum_{j=0}^{\infty} b_j \cdot t^j \;:=\; \sum_{k=0}^{\infty} \left(\sum_{i+j=k} a_i \cdot b_j \right) \cdot t^k \in R[[t]].$$

Note that by this definition we get immediately that

$$\sum_{i=0}^{\infty} a_i \cdot t^i \;=\; \sum_{i=0}^{\infty} b_i \cdot t^i \quad\Longleftrightarrow\quad a_i = b_i \;\forall\, i \in \mathbb{N}.$$

If $a_i = 0$ for $i \geq n$ then we write for short

$$\sum_{i=0}^{n} a_i \cdot t^i = \sum_{i=0}^{\infty} a_i \cdot t^i.$$

Remark 6.7
The definition of the multiplication in $R[[t]]$ comes from the wish to get a generalised distributive law for such *infinite sums*:

$$\left(\sum_{i=0}^{\infty} a_i \cdot t^i \right) \cdot \left(\sum_{j=0}^{\infty} b_j \cdot t^j \right) = \sum_{i=0}^{\infty} \left(a_i \cdot t^i \cdot \left(\sum_{j=0}^{\infty} b_j \cdot t^j \right) \right) =$$

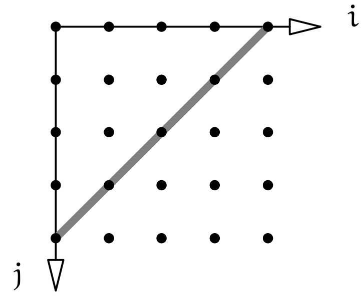
$$\sum_{i=0}^{\infty} \left(\left(\sum_{j=0}^{\infty} a_i \cdot t^i \cdot b_j \cdot t^j \right) \right) = \sum_{i=0}^{\infty} \left(\left(\sum_{j=0}^{\infty} a_i \cdot b_j \cdot t^{i+j} \right) \right). \tag{25}$$

If the operations which we just did in $R[[t]]$ are correct, i.e. if all equations are correct, then our notation of the elements of $R[[t]]$ as infinite sums is a useful notation. However, the expression on the right hand side in (25) is in the above form not yet recognisable as a formal power series, i.e. as element of $R[[t]]$. Many of the t^{i+j} coincide for different values of i and j; e.g. we get t^2 by $(i, j) = (2, 0)$ and by $(i,j) = (1,1)$ and by $(i,j) = (0,2)$. The coefficient of t^2 in the power series in (25) should therefore be the sum of the $a_i \cdot b_j$ for these (i,j), i.e. it should be $a_2 \cdot b_0 + a_1 \cdot b_1 + a_0 \cdot b_2$.

The main question is if for each k there are only finitely many pairs (i,j) such that $i+j=k$? The answer to this question is yes! To see this note that by assumption i and j are non-negative. If we then fix k the choice of i determines j as $j=k-i$. But since i can only take values between 0 and k we have a finite number of pairs (i,j) with $i+j=k$, namely the $k+1$ pairs:

$$(k,0),\ \ (k-1,1),\ \ (k-2,2),\ \ \ldots\ ,\ \ (1,k-1),\ \ (0,k).$$

One can visualise the pairs (i,j) whose sum is k as points in a coordinate system with axes labeled by i and j:



Exactly the pairs (i,j) on the diagonal from $(k,0)$ to $(0,k)$ have the property that the sum $i+j$ has the value k.

The coefficient of t^k in the right hand side of (25) must therefor have the form

$$\sum_{i+j=k} a_i \cdot b_j = \sum_{i=0}^{k} a_i \cdot b_{k-i}.$$

Theorem 6.8
If R is a commutative ring with one then the ring of formal power series $(R[[t]],+,\cdot)$ is a commutative ring with one $1_{R[[t]]} = t^0$.

Proof: By definition $+$ and $\cdot$ are two binary operations on $R[[t]]$. Let $\sum_{i=0}^{\infty} a_i \cdot t^i, \sum_{i=0}^{\infty} b_i \cdot t^i, \sum_{i=0}^{\infty} c_i \cdot t^i \in R[[t]]$ be given. Then the associativity of the addition in R gives

$$\begin{aligned}\left(\sum_{i=0}^{\infty} a_i \cdot t^i + \sum_{i=0}^{\infty} b_i \cdot t^i\right) + \sum_{i=0}^{\infty} c_i \cdot t^i &= \sum_{i=0}^{\infty} \big((a_i+b_i)+c_i\big) \cdot t^i \\ &= \sum_{i=0}^{\infty} \big(a_i+(b_i+c_i)\big) \cdot t^i = \sum_{i=0}^{\infty} a_i \cdot t^i + \left(\sum_{i=0}^{\infty} b_i \cdot t^i + \sum_{i=0}^{\infty} c_i \cdot t^i\right)\end{aligned}$$

and the commutativity of the addition in R implies

$$\sum_{i=0}^{\infty} a_i \cdot t^i + \sum_{i=0}^{\infty} b_i \cdot t^i = \sum_{i=0}^{\infty} (a_i + b_i) \cdot t^i$$
$$= \sum_{i=0}^{\infty} (b_i + a_i) \cdot t^i = \sum_{i=0}^{\infty} b_i \cdot t^i + \sum_{i=0}^{\infty} a_i \cdot t^i.$$

Moreover, the constant map $0_{R[[t]]} = \sum_{i=0}^{\infty} 0 \cdot t^i$ satisfies

$$0_{R[[t]]} + \sum_{i=0}^{\infty} a_i \cdot t^i = \sum_{i=0}^{\infty} (0 + a_i) \cdot t^i = \sum_{i=0}^{\infty} a_i \cdot t^i,$$

and for $\sum_{i=0}^{\infty} (-a_i) \cdot t^i \in R[[t]]$ we get

$$\sum_{i=0}^{\infty} (-a_i) \cdot t^i + \sum_{i=0}^{\infty} a_i \cdot t^i = \sum_{i=0}^{\infty} (-a_i + a_i) \cdot t^i = 0_{R[[t]]},$$

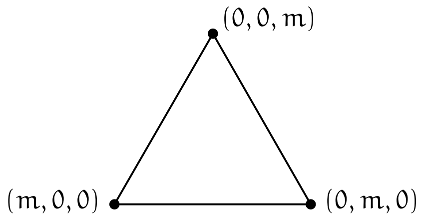
so that $(R[[t]], +)$ is an abelian group with the constant zero-map as neutral element.

Note that

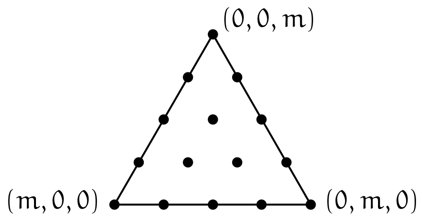
$$\sum_{k+l=m} \left(\sum_{i+j=k} a_i \cdot b_j \right) \cdot c_l = \sum_{i+j+l=m} a_i \cdot b_j \cdot c_l = \sum_{i+k=m} \left(a_i \cdot \sum_{j+l=k} b_j \cdot c_l \right), \tag{26}$$

since in these sums triple (i, j, l) of natural numbers with the property $i + j + l = m$ occurs exactly once[25] and since in R the associativity law for the multiplication holds

[25] We can visualise this in a graphical way by considering (i, j, l) as coordinates in three-space $\mathbb{R}^3$. The equation $i + j + l = m$ for a fixed m then describes a plane in this space, namely the plane spanned by the three points $(m, 0, 0)$, $(0, m, 0)$ and $(0, 0, m)$. If we join these three points to each other in this plane we get a triangle:



The points with integer coordinates in this triangle are precisely the triples of non-negative integers whose sum is m:



In the left hand side of (26) the set of these points is partitioned as follows:

$$\bigcup_{l=0}^{m} \bigcup_{i+j=m-l} \{(i, j, l)\}.$$

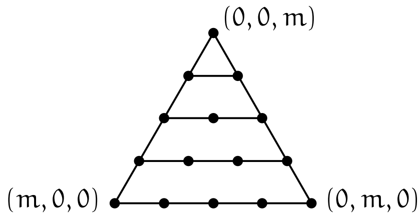
In the inner sum we collect just those integer triples (i, j, l) in the triangle for which the coordinate l is constant and for which $i + j = m - l$, i.e. the points lie on a line parallel to the line through $(m, 0, 0)$ and $(0, m, 0)$:

as well as the distributivity law. We therefore get

$$\begin{aligned}
\left(\sum_{i=0}^{\infty} a_i \cdot t^i \cdot \sum_{j=0}^{\infty} b_j \cdot t^j\right) \cdot \sum_{l=0}^{\infty} c_l \cdot t^l &= \sum_{k=0}^{\infty}\left(\sum_{i+j=k} a_i \cdot b_j\right) \cdot t^k \cdot \sum_{l=0}^{\infty} c_l \cdot t^l \\
&= \sum_{m=0}^{\infty}\left(\sum_{k+l=m}\left(\sum_{i+j=k} a_i \cdot b_j\right) \cdot c_l\right) \cdot t^m \\
&= \sum_{m=0}^{\infty}\left(\sum_{i+k=m} a_i \cdot \sum_{j+l=k} b_j \cdot c_l\right) \cdot t^m \\
&= \sum_{i=0}^{\infty} a_i \cdot t^i \cdot \sum_{k=0}^{\infty}\left(\sum_{j+l=k} b_j \cdot c_l\right) \cdot t^k \\
&= \sum_{i=0}^{\infty} a_i \cdot t^i \cdot \left(\sum_{j=0}^{\infty} b_j \cdot t^j \cdot \sum_{l=0}^{\infty} c_l \cdot t^l\right),
\end{aligned}$$

so that the multiplication on $R[[t]]$ is associative. Moreover, the commutativity of the multiplication on R gives

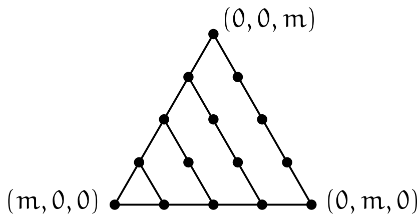
$$\begin{aligned}
\sum_{i=0}^{\infty} a_i \cdot t^i \cdot \sum_{j=0}^{\infty} b_j \cdot t^j &= \sum_{k=0}^{\infty}\left(\sum_{i+j=k} a_i \cdot b_j\right) \cdot t^k \\
&= \sum_{k=0}^{\infty}\left(\sum_{j+i=k} b_j \cdot a_i\right) \cdot t^k = \sum_{j=0}^{\infty} b_j \cdot t^j \cdot \sum_{i=0}^{\infty} a_i \cdot t^i.
\end{aligned}$$



In the right hand side of (26) we partition the set as follows:

$$\bigcup_{i=0}^{m} \bigcup_{j+l=m-i} \{(i, j, l)\}.$$

In the inner sum all those integer triples (i, j, l) in the triangle are collected for which the coordinate i is constant and for which $j + l = m - i$, i.e. the points lie on a line parallel to the line through $(0, m, 0)$ and $(0, 0, m)$:



In both cases each of the integer triples (i, j, l) in the triangle is considered exactly once.

And finally for $1_{R[[t]]} = t^0 = \sum_{j=0}^{\infty} e_j \cdot t^j$ with $e_0 = 1$ and $e_j = 0$ for $j \geq 1$ we get

$$1_{R[[t]]} \cdot \sum_{i=0}^{\infty} a_i \cdot t^i = \sum_{k=0}^{\infty} \left(\sum_{j+i=k} e_j \cdot a_i \right) \cdot t^k = \sum_{k=0}^{\infty} a_k \cdot t^k,$$

so that using the commutativity of the multiplication in $R[[t]]$ we see that t^0 is the 1-element of $R[[t]]$.

It remains to show the distributivity law:

$$\begin{aligned}
\sum_{i=0}^{\infty} a_i \cdot t^i \cdot \left(\sum_{j=0}^{\infty} b_j \cdot t^j + \sum_{j=0}^{\infty} c_j \cdot t^j \right) &= \sum_{i=0}^{\infty} a_i \cdot t^i \cdot \sum_{j=0}^{\infty} (b_j + c_j) \cdot t^j \\
&= \sum_{k=0}^{\infty} \left(\sum_{i+j=k} a_i \cdot (b_j + c_j) \right) \cdot t^k \\
&= \sum_{k=0}^{\infty} \left(\sum_{i+j=k} (a_i \cdot b_j + a_i \cdot c_j) \right) \cdot t^k \\
&= \sum_{k=0}^{\infty} \left(\sum_{i+j=k} a_i \cdot b_j \right) \cdot t^k + \sum_{k=0}^{\infty} \left(\sum_{i+j=k} a_i \cdot c_j \right) \cdot t^k \\
&= \sum_{i=0}^{\infty} a_i \cdot t^i \cdot \sum_{j=0}^{\infty} b_j \cdot t^j + \sum_{i=0}^{\infty} a_i \cdot t^i \cdot \sum_{j=0}^{\infty} c_j \cdot t^j.
\end{aligned}$$

The second distributivity law follows right away by the commutativity of the multiplication.

Altogether we have shown that $(R[[t]], +, \cdot)$ is a commutative ring with one. $\square$

Remark 6.9

Our definition of a formal power series sometimes confuses beginners. In analysis a power series is a map on some interval $(a - \varepsilon, a + \varepsilon)$ in the real numbers, and one evaluates such a map by replacing the indeterminate t by a real number from the interval $(a - \varepsilon, a + \varepsilon)$. Replacing t by a real number only makes sense if one has the notion of *convergence* of sequences of real numbers at hand. This notion cannot easily be generalised to arbitrary rings. We therefore should not try to replace t by any values in R, except maybe by 0!

Yet, we still defined a power series as a scheme which represents a map, namely as a map from $\mathbb{N}$ into the ring R. A natural way to represent such a map would be by a value table:

k	0	1	2	3	$\ldots$
$f(k)$	a_0	a_1	a_2	a_3	$\ldots$

In principle, the power series $\sum_{k=0}^{\infty} a_k \cdot t^k$ is just a compact way of writing such a value table. The column

3
a_3

is replaced by the term $a_3 \cdot t^3$, and the "$\cdot t^3$" is just there to recall that the fourth column of the value table is represented. Well, and instead of separating the different columns of the table by a vertical bar we connect them by the "+"-symbol:

$$a_0 \cdot t^0 + a_1 \cdot t^1 + a_2 \cdot t^2 + a_3 \cdot t^3 + \ldots.$$

If $f = \sum_{k=0}^{\infty} a_k \cdot t^k \in R[[t]]$ is a power series then $f(k) = a_k$ by definition.

Why do we prefer this notation?

We have introduced an addition and a multiplication of power series, i.e. of maps from $\mathbb{N}$ to R. These can be expressed in a natural way using this notation. That's all there is to it!

In order to be able to compute in the ring of power series it suffices to know the rules in Definition 6.6. We then can use $f = \sum_{k=0}^{\infty} a_k \cdot t^k$ as a formal expression without any further meaning which can be manipulated by the given rules. Just forget that the expression represents a map, as long as you recall the rules. $\square$

Exercise 6.10
Let R be a commutative ring with one and $f = \sum_{k=0}^{\infty} a_k \cdot t^k \in R[[t]]$ a formal power series over R. Show that f is a unit in $R[[t]]$ if and only if a_0 is a unit in R.

Hint, if a_0 is a unit in R then we are looking for a series $g = \sum_{k=0}^{\infty} b_k \cdot t^k$ such that $f \cdot g = t^0$. Expanding the left hand side of the last equation leads to a bunch of equations for the coefficients of the power series, and these can be solved recursively.

B) **Subrings**

Definition 6.11
Let R a ring with one and $S \subseteq R$. S is called a *subring* of R if

a. $1_R \in S$,

b. $a + b \in S$ for all $a, b \in S$,

c. $-a \in S$ for all $a \in S$, and

d. $a \cdot b \in S$ for all $a, b \in S$.

If R a field and S a subring of R such that additionally $a^{-1} \in S$ for all $a \in S \setminus \{0\}$ then S is called a *subfield* of R.

Note that a subring S of R is in particular a ring with respect to the restriction of the addition and multiplication of R to S. The analogous statement for fields holds true as well.

Example 6.12
$\mathbb{Z}$ is a subring of $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. $\mathbb{Q}$ is a subfield of $\mathbb{R}$ and $\mathbb{C}$. $\mathbb{R}$ is a subfield of $\mathbb{C}$.

The most important example of a ring besides the integers is the polynomial ring which we obtain as a subring of the power series ring.

Definition 6.13
If R is a commutative ring then

$$\begin{aligned} R[t] &:= \left\{ \sum_{k=0}^{\infty} a_k \cdot t^k \in R[[t]] \;\middle|\; \text{only finitely many } a_k \text{ are non-zero} \right\} \\ &= \left\{ \sum_{k=0}^{n} a_k \cdot t^k \in R[[t]] \;\middle|\; n \in \mathbb{N}, a_0, \ldots, a_n \in R \right\} \end{aligned}$$

is called the *polynomial ring* over R in the indeterminate t and the elements of $R[t]$ are called *polynomials*. For $0 \neq f = \sum_{k=0}^{\infty} a_k \cdot t^k \in R[t]$ we call

$$\deg(f) = \max\{k \mid a_k \neq 0\}$$

the *degree* of the *polynomial* f and $\mathrm{lc}(f) := a_{\deg(f)}$ its *leading coefficient*. Moreover, we set $\deg(0) = -\infty$ and $\mathrm{lc}(0) := 0$.

Note that by convention in Definition 6.6 each polynomial in $R[t]$ has the form

$$\sum_{k=0}^{n} a_k \cdot t^k$$

for some $n \in \mathbb{N}$.

Example 6.14
$3 \cdot t^4 - t^2 + 5 \cdot t^0 \in \mathbb{Z}[t]$ is a polynomial of degree $\deg(f) = 4$ and with leading coefficient $\mathrm{lc}(f) = 3$.

Since $R[t]$ is closed with respect to addition, negatives and multiplication and since $1_{R[[t]]} = t^0 \in R[t]$ we get the following theorem.

Theorem 6.15
If R *is a commutative ring with one then* $R[t]$ *is a subring of* $R[[t]]$*. In particular,* $R[t]$ *is a commutative ring with one.*

Proof: Let $f = \sum_{k=0}^{m} a_k \cdot t^k$, $g = \sum_{k=0}^{n} b_k \cdot t^k \in R[t]$ be given (where we allow that all a_k or b_k are zero) then

$$f + g = \sum_{k=0}^{\max\{m,n\}} (a_k + b_k) \cdot t^k \in R[t], \tag{27}$$

$$-f = \sum_{k=0}^{m} (-a_k) \cdot t^k \in R[t]$$

and

$$f \cdot g = \sum_{k=0}^{m+n} \left(\sum_{i=0}^{k} a_i \cdot b_{k-i} \right) \cdot t^k \in R[t], \tag{28}$$

where we use the convention that $a_k = 0$ for $k > m$ and $b_k = 0$ for $k > n$. In order to see that in (28) no term of degree larger than $n + m$ occurs one should note that the coefficient of t^k for $k > n + m$ has the form

$$\sum_{i=0}^{m} a_i \cdot b_{k-i} + \sum_{i=m+1}^{k} a_i \cdot b_{k-i}.$$

The second sum is zero since all a_i are so, while the first sum is zero since all b_{k-i} are so. □

The following degree formulae for polynomials follow right away from the above proof. We here use the convention that $m + -\infty = -\infty$ and $\max\{m, -\infty\} = m$ for all $m \in \mathbb{N} \cup \{-\infty\}$.

Proposition 6.16 (Degree Formulae)
Let R be a commutative ring with one and $f, g \in R[t]$ be two polynomials. Then

a. $\deg(f + g) \leq \max\{\deg(f), \deg(g)\}$.

b. $\deg(f \cdot g) \leq \deg(f) + \deg(g)$.

c. $\deg(f \cdot g) = \deg(f) + \deg(g)$ *if and only if* $\mathrm{lc}(f) \cdot \mathrm{lc}(g) \neq 0$.
In this case we also have $\mathrm{lc}(f \cdot g) = \mathrm{lc}(f) \cdot \mathrm{lc}(g)$.

Proof: If $f = 0$ or $g = 0$ then the statements are obviously correct. We thus may assume without loss of generality that $f \neq 0 \neq g$. Then a. follows right away from (27) and b. from (28). For c. note that in (28) the coefficient of of t^{m+n} is just $a_m \cdot b_n = \mathrm{lc}(f) \cdot \mathrm{lc}(g)$. □

Exercise 6.17 a. Let R be a commutative ring with one and $S \subset R$ a non-empty subset such that

- $x + y \in S$ for all $x, y \in S$,
- $-x \in S$ for all $x \in S$,
- $x \cdot y \in S$ for all $x, y \in S$ and
- $1_R \in S$.

Show S is a commutative ring with one with respect to the restriction of the addition and the multiplication of R to S.

b. Show $\mathbb{Z}[i] := \{a + i \cdot b \in \mathbb{C} \mid a, b \in \mathbb{Z}\}$ is a commutative ring with one, where the addition and the multiplication are just the addition and multiplication of complex numbers. We call this ring the *ring of Gaussian integers.*

c. Compute the group of units $\mathbb{Z}[i]^*$ of the ring $\mathbb{Z}[i]$.

Exercise 6.18
For $\omega \in \mathbb{Z}$, $\omega \geq 2$ we denote by $\sqrt{-\omega}$ the complex number $i \cdot \sqrt{\omega}$.

a. Show that $\mathbb{Z}\left[\sqrt{-\omega}\right] := \{a + b \cdot \sqrt{-\omega} \in \mathbb{C} \mid a, b \in \mathbb{Z}\}$ is a commutative ring with one, where the addition and the multiplication is the addition and multiplication of complex numbers.

b. Show $\mathbb{Z}\left[\sqrt{-\omega}\right]^* = \{1, -1\}$.

C) **Ring Homomorphisms**

With any new structure we also define the maps which respect the structure. Note that in a ring with one besides the addition and the multiplication also the existence of a 1-element is part of the structure. We will therefore require that a structure preserving map respects the addition, the multiplication and the 1-element.

Definition 6.19
Let R and S be two rings with one. A map $\varphi : R \to S$ is called a *ring homomorphism* if

a. $\varphi(a+b) = \varphi(a) + \varphi(b)$ for all $a, b \in R$,

b. $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$ for all $a, b \in R$ and

c. $\varphi(1_R) = 1_S$.

If φ a ring homomorphism the we call φ

- a *monomorphism* if φ is injective;
- an *epimorphism* if φ is surjective;
- an *isomorphism* if φ is bijective.

We say that two rings R and S are *isomorphic*, if there exists an isomorphism from R to S. We then write $R \cong S$ for short.

Example 6.20
If $S \subseteq R$ is a subring of the ring R then the canonical inclusion $i_S : S \to R$ is a ring homomorphism.

Lemma 6.21
If $\varphi : R \to S$ is a bijective ring homomorphism then also $\varphi^{-1} : S \to R$ is a ring homomorphism.

Proof: That φ^{-1} respects the addition follows from Proposition 1.51 c. since φ is a homomorphism of the abelian groups from $(R,+)$ to $(S,+)$. That it respects the multiplication works with the same proof, and it respects the 1-element since $\varphi(1_R) = 1_S$. □

Lemma 6.22
If $\varphi : R \longrightarrow S$ is a ring homomorphism then $\mathrm{Im}(\varphi)$ *is a subring of* S.

Proof: By Proposition 1.51 $\mathrm{Im}(\varphi)$ is a subgroup of $(S,+)$ so that $\mathrm{Im}(\varphi)$ is closed with respect to addition and negatives. Moreover,

$$1_S = \varphi(1_R) \in \mathrm{Im}(\varphi)$$

and for $\varphi(a), \varphi(b) \in \mathrm{Im}(\varphi)$ we have

$$\varphi(a) \cdot \varphi(b) = \varphi(a \cdot b) \in \mathrm{Im}(\varphi).$$

□

Since a ring homomorphism $\varphi : R \longrightarrow S$ is by definition a group homomorphism from $(R, +)$ to $(S, +)$ we get the following criterion for injectivity for ring homomorphisms from Lemma 1.52 for free. Note that $\mathrm{Ker}(\varphi) = \{a \in R \mid \varphi(a) = 0_S\}$.

Lemma 6.23
A ring homomorphism $\varphi : R \longrightarrow S$ *is a monomorphism if and only if* $\mathrm{Ker}(\varphi) = \{0_R\}$.

Remark 6.24
If R is a commutative ring with one then the map

$$\iota : R \longrightarrow R[t] : a \mapsto a \cdot t^0$$

is a monomorphism of rings and thus R is isomorphic to the subring $\mathrm{Im}(\iota) = \{a \cdot t^0 \mid a \in R\}$ of $R[[t]]$. We will use this isomorphism from now on so as not to distinguish any more between the elements of R and the constant polynomials, i.e. we write e.g. $2t^2 + 3$ instead of $2t^2 + 3t^0$.

Exercise 6.25
Let K be a field, R be a commutative ring with $1_R \neq 0_R$ and $\varphi : K \longrightarrow R$ be a ring homomorphism. Show φ is a monomorphism.

Exercise 6.26
Let S be a commutative ring with one, $R \subseteq S$ a subring and $b \in S$.

a. We define
$$f(b) = \sum_{k=0}^{n} a_k \cdot b^k \in S$$
for $f = \sum_{k=0}^{n} a_k \cdot t^k \in R[t]$. Show that the map
$$\varphi_b : R[t] \longrightarrow S : f \mapsto f(b)$$
is a ring homomorphism. We call φ_b a *substitution morphism*.

b. If b is a zero of the polynomial $g = t^n + \alpha_{n-1} \cdot t^{n-1} + \ldots + \alpha_1 \cdot t + \alpha_0 \in R[t]$ then
$$\mathrm{Im}(\varphi_b) = \left\{a_0 + a_1 \cdot b + a_2 \cdot b^2 + \ldots + a_{n-1} \cdot b^{n-1} \;\middle|\; a_0, \ldots, a_{n-1} \in R\right\}.$$
We denote this subring of S by $R[b] = \mathrm{Im}(\varphi_b)$.

D) **Ideals**

If R is a commutative ring with one and S is a subring of R then in particular $(S, +)$ a normal subgroup of the abelian group $(R, +)$. Therefore the quotient group $(R/S, +)$ is defined where for two cosets $\overline{a}, \overline{b} \in R/S$ the sum is defined as $\overline{a} + \overline{b} = \overline{a + b}$. We would like to carry also the second operation that we have on R and S over to the quotient group R/S by $\overline{a} \cdot \overline{b} = \overline{a \cdot b}$ and we would like to have that then R/S is a commutative ring with one. However, this does not work! The 0-element of R/S is necessarily $\overline{0}$ and thus for an arbitrary $\overline{a} \in R/S$ and $b \in S$ we would like to have

$$S = \overline{0} = \overline{a \cdot 0} = \overline{a} \cdot \overline{b} = \overline{a \cdot b} = (a \cdot b) + S.$$

Thus $a \cdot b \in S$, i.e. S would be closed with respect to the multiplication by arbitrary element of R. Since by assumption $1_R \in S$ this would imply that for each element $a \in R$ we have

$$a = a \cdot 1_R \in S,$$

and hence $S = R$. I.e. the only subring for which the corresponding quotient group $(R/S, +)$ can be given the structure of a ring in this way would be the ring R itself. But then the ring R/S only contains one element and is boring.

If we nevertheless want to build quotient structures then we have to replace the notion of a subring by a more suitable notion. We have already seen that we need a structure which on the one hand is a subgroup of $(R, +)$ and on the other hand is closed with respect to the multiplication by an arbitrary element of R. This leads to the following definition, and up to the end of this section we restrict our attention to rings which are *commutative.*

Definition 6.27
Let R be a commutative ring with one and $\emptyset \neq I \subseteq R$ be a non-empty subset of R. I is called an *ideal* of R if

(1) $a + b \in I$ for all $a, b \in I$ and

(2) $r \cdot a \in I$ for all $r \in R$ and $a \in I$.

We then write $I \trianglelefteq R$ since ideals are the analoga of normal subgroups in the case of rings.

Remark 6.28
Let R be a commutative ring with one and $I \trianglelefteq R$. Then $(I, +)$ is a subgroup of $(R, +)$. This follows right away from the subgroup criterion in Proposition 1.28 since $-1_R \in R$ and for $a \in I$ thus also $-a = -1_R \cdot a \in I$.

Note that by the definition of an ideal and by induction we get

$$\sum_{k=1}^{n} r_k \cdot a_k \in I$$

for all $r_k \in R$ and $a_k \in I$.

Example 6.29

a. If R is a commutative ring with one then $\{0_R\}$ and R are the *trivial* ideals of R.

b. $n\mathbb{Z}$ is an ideal in $\mathbb{Z}$ for each $n \in \mathbb{Z}$ since the set is closed with respect to addition and with respect to multiplication by arbitrary integers.

Since each ideal of $(\mathbb{Z}, +, \cdot)$ is in particular a subgroup of $(\mathbb{Z}, +)$ by the above example, the classification of the subgroups of $(\mathbb{Z}, +)$ in Proposition 1.39 implies the following corollary.

Corollary 6.30
For a subset $U \subseteq \mathbb{Z}$ the following statements are equivalent:

a. U *is an ideal of* $(\mathbb{Z},+,\cdot)$.

b. U *is a subgroup of* $(\mathbb{Z},+)$.

c. $\mathsf{U}=n\mathbb{Z}$ *for some integer* $n\geq 0$.

As for subgroups we would like to know how ideals behave with respect to settheoretic operations.

Proposition 6.31
If R *is a commutative ring with one and* $I_\lambda \trianglelefteq R$ *are ideals for* $\lambda\in\Lambda$ *then also* $\bigcap_{\lambda\in\Lambda} I_\lambda \trianglelefteq R$ *is an ideal.*

Proof: Let $a,b\in\bigcap_{\lambda\in\Lambda} I_\lambda$ and $r\in R$. Then $a+b\in I_\lambda$ and $r\cdot a\in I_\lambda$ since I_λ is an ideal. Thus also

$$a+b,\ r\cdot a\in\bigcap_{\lambda\in\Lambda} I_\lambda.$$

Moreover, $0_R\in I_\lambda$ since $(I_\lambda,+)$ is a subgroup of $(R,+)$ and hence $0_R\in\bigcap_{\lambda\in\Lambda} I_\lambda$ so that the set is non-empty. □

Definition 6.32
Let R be a commutative ring and $M\subseteq R$ a subset. We define the *ideal generated* by M as

$$\langle M\rangle_R=\bigcap_{M\subseteq I\trianglelefteq R} I,$$

the intersection of all ideals in R which contain M.

Proposition 6.33
Let R *be a commutative ring with one and* $\emptyset\neq M\subseteq R$. *Then the ideal generated by* M *is*

$$\langle M\rangle_R=\left\{\sum_{k=1}^{n} r_k\cdot a_k\ \middle|\ a_k\in M, r_k\in R, n\geq 1\right\}\trianglelefteq R$$

the set of all finite linear combinations *of elements in* M *with coefficients in* R.

Proof: We set

$$J=\left\{\sum_{k=1}^{n} r_k\cdot a_k \mid a_k\in M, r_k\in R, n\geq 1\right\}$$

and shown first that J is an ideal of R.

Since M is a non-empty set so is J. Let $\sum_{k=1}^{n} r_k\cdot a_k, \sum_{k=1}^{m} s_k\cdot b_k\in J$ with $r_k,s_k\in R$ and $a_k,b_k\in M$ we then define $r_k=s_{k-n}$ and $a_k=b_{k-n}$ for $k=n+1,\ldots,n+m$ and get thus

$$\sum_{k=1}^{n} r_k\cdot a_k+\sum_{k=1}^{m} s_k\cdot b_k=\sum_{k=1}^{n+m} r_k\cdot a_k\in J.$$

Since moreover for $r\in R$ also $r\cdot r_k\in R$ we get

$$r\cdot\sum_{k=1}^{n} r_k\cdot a_k=\sum_{k=1}^{n}(r\cdot r_k)\cdot a_k\in J.$$

Hence J is an ideal in R.

Moreover, $M \subset J$ since $a = 1_R \cdot a \in J$ for all $a \in M$. Thus by the definition we get

$$\langle M \rangle_R = \bigcap_{M \subseteq I \trianglelefteq R} I \subseteq J.$$

On the other hand for $\sum_{k=1}^{n} r_k \cdot a_k \in J$ with $r_k \in R$ and $a_k \in M$ we have that

$$\sum_{k=1}^{n} r_k \cdot a_k \in I$$

for each ideal $I \trianglelefteq R$ which contains M by Remark 6.28. Therefore we get

$$J \subseteq \bigcap_{M \subseteq I \trianglelefteq R} I = \langle M \rangle_R.$$

□

Example 6.34
If R is a commutative ring with one and $a, b \in R$ then

$$\langle a \rangle_R = \{r \cdot a \mid r \in R\}$$

and

$$\langle a, b \rangle_R = \{r \cdot a + s \cdot b \mid r, s \in R\}.$$

In particular, $n\mathbb{Z} = \langle n \rangle_{\mathbb{Z}}$.

Exercise 6.35
Show that

$$I = \{f \in \mathbb{Z}[t] \mid f(5) = 0\}$$

is an ideal in $\mathbb{Z}[t]$.

Exercise 6.36
Let R be a commutative ring. R is a field if and only if R has exactly two ideals.

Exercise 6.37
Let R be a ring and let $I_k \trianglelefteq R$, $k \in \mathbb{N}$, be ideals with the property

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq I_3 \subseteq \ldots,$$

i.e. $I_k \subseteq I_{k+1}$ for all $k \in \mathbb{N}$. Show that then

$$\bigcup_{k \in \mathbb{N}} I_k \trianglelefteq R$$

is an ideal in R.

E) **Quotient Rings**

Theorem 6.38
Let R *be a commutative ring with one and* $I \trianglelefteq R$ *be an ideal. Then by*

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

for $\overline{a}, \overline{b} \in R/I$ *a binary operation is defined on the abelian group* $(R/I, +)$*, and* $(R/I, +, \cdot)$ *is a commutative ring with* 1*-element* $1_{R/I} = \overline{1_R}$*. We call* R/I *the* quotient ring *of* R *by* I.

Proof: We first have to show that the operation is well defined, i.e. that it does not depend on the choice of the representative of the coset. Let therefore $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$. Then by definition $a = a' + c$ and $b = b' + d$ for some $c, d \in I$, and we get

$$a \cdot b = (a' + c) \cdot (b' + d) = a' \cdot b' + \big(a' \cdot d + c \cdot b' + c \cdot d\big)$$

with $a' \cdot d + c \cdot b' + c \cdot d \in I$. Hence,

$$\overline{a \cdot b} = a \cdot b + I = a' \cdot b' + I = \overline{a' \cdot b'},$$

and the multiplication is well defined. The associativity and the commutativity of the multiplication are induced by the corresponding properties of the multiplication on R. Moreover, $\overline{1} \cdot \overline{a} = \overline{1 \cdot a} = \overline{a}$ for all $\overline{a} \in R/I$ so that $(R/I, +, \cdot)$ is indeed a commutative ring with one $\overline{1}$. $\square$

Example 6.39
$(Z_n, +, \cdot)$ is a commutative ring with one by

$$\overline{a} + \overline{b} = \overline{a + b} \quad \text{and} \quad \overline{a} \cdot \overline{b} = \overline{a \cdot b}$$

for all $a, b \in \mathbb{Z}$ and $n \geq 0$.

Note, in $\mathbb{Z}_2 = \{\overline{0}, \overline{1}\}$ is obviously every element apart from $\overline{0}$ a unit, so that $(Z_2, +, \cdot)$ is a field. Since, moreover, any field K contains at least two elements, namely $0_K \neq 1_K$, $\mathbb{Z}_2$ is the smallest possible field.

Exercise 6.40
Which of the following rings is a field?

a. $\mathbb{Z}_4$.

b. $\mathbb{Z}_7$.

Exercise 6.41
For a positive integer n we define the map

$$\phi_n : \mathbb{Z}[t] \longrightarrow \mathbb{Z}_n[t] : \sum_{k=0}^{n} a_k \cdot t^k \mapsto \sum_{k=0}^{n} \overline{a_k} \cdot t^k.$$

Show that ϕ_n is a ring epimorphism. We call ϕ_n the *reduction modulo* n.

Exercise 6.42
a. Compute $\mathbb{Z}_6^*$.

b. Compute $\mathbb{Z}_8^*$.

c. Compute $\mathbb{Z}_{15}^*$.

d. Formulate a conjecture under which circumstances an element $\overline{z} \in \mathbb{Z}_n$ for $n \geq 2$ is a unit.

e. Show $\overline{n-1} \in \mathbb{Z}_n$ is a unit for all $n \geq 2$.

F) **Homomorphism Theorem**

Theorem 6.43 (Homomorphism Theorem)
Let $\varphi : R \longrightarrow S$ *be a ring homomorphism of commutative rings with one. Then* $\mathrm{Ker}(\varphi) = \{a \in R \mid \varphi(a) = 0\} \trianglelefteq R$ *is an ideal and*

$$\overline{\varphi} : R/\mathrm{Ker}(\varphi) \longrightarrow \mathrm{Im}(\varphi)$$

is an isomorphism. In particular, $R/\mathrm{Ker}(\varphi) \cong \mathrm{Im}(\varphi)$.

Proof: By Proposition 1.51 $(\mathrm{Ker}(\alpha), +)$ is a subgroup of $(R, +)$. In particular, it is non-empty and closed with respect to the addition. Let $a \in \mathrm{Ker}(\alpha)$ and $r \in R$. Then

$$\varphi(r \cdot a) = \varphi(r) \cdot \varphi(a) = \varphi(r) \cdot 0 = 0.$$

Hence, $r \cdot a \in \mathrm{Ker}(\alpha)$ and $\mathrm{Ker}(\alpha)$ is an ideal. The Homomorphism Theorem 4.50 for groups then implies that $\overline{\varphi}$ is an isomorphism of abelian groups. Since moreover

$$\overline{\varphi}\big(\overline{a} \cdot \overline{b}\big) = \overline{\varphi}\big(\overline{a \cdot b}\big) = \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b) = \overline{\varphi}\big(\overline{a}\big) \cdot \overline{\varphi}\big(\overline{b}\big)$$

and $\overline{\varphi}\big(\overline{1}\big) = \varphi(1) = 1$ the map $\overline{\varphi}$ is indeed an isomorphism of rings. □

Exercise 6.44

a. Find all ring homomorphisms $\varphi : \mathbb{Z} \longrightarrow \mathbb{Z}_6$.

b. Find all ring homomorphisms $\varphi : \mathbb{Z}_6 \longrightarrow \mathbb{Z}$.

c. Find all ring homomorphisms $\varphi : \mathbb{Q} \longrightarrow \mathbb{R}$.

7 Divisibility in Rings

In the previous section we have introduced the notion of commutative rings with one. The model for this notion were the integers, and they are yet again the model for the properties of rings which we study in this section.

A) Integral Domains

The central notion of this subsection will be the divisibility. A well known property of the integers is that the product of two integers can only be zero if one of them is zero. This does not hold any more in more general rings. If we consider for instance the ring $\mathbb{Z}_4$ then the coset $\overline{2} \neq \overline{0}$ while $\overline{2}\cdot\overline{2} = \overline{4} = \overline{0}$. This has unpleasant consequences since $\overline{0}$ can be written in several ways as a multiple of $\overline{2}$:

$$\overline{2} \cdot \overline{2} = \overline{0} = \overline{0} \cdot \overline{2}.$$

In such a ring obviously the cancellation rules for the multiplication do not hold. However, these are vital for the notion of divisibility. We therefore introduce a name for rings which behave well in that respect.

Definition 7.1
Let R be a commutative ring with one and $a \in R$.

a. a is called a *zero divisor* if there exists a $0 \neq b \in R$ such that $a \cdot b = 0$.

b. R is called an *integral domain* if 0 is the unique zero divisor in R.

Example 7.2

a. If R is not the nullring then 0 is a zero divisor, since $0 \cdot 1 = 0$ and $1 \neq 0$.

b. If $a \in R^*$ a unit then a is not a zero divisor.
Since a has an inverse $a^{-1} \in R$ we deduce from $a \cdot b = 0$ that

$$b = 1 \cdot b = \left(a^{-1} \cdot a\right) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0.$$

c. From b. we deduce that each field is an integral domain, since 0 is the only element which is not a unit. In particular, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are integral domains.

d. Each subring of an integral domain is an integral domain. In particular, $\mathbb{Z}$ and

$$\mathbb{Z}\left[\sqrt{-5}\right] = \left\{a + b \cdot \sqrt{-5} \in \mathbb{C} \;\middle|\; a, b \in \mathbb{Z}\right\}$$

are integral domains.

e. If R is an integral domain then $R[t]$ is an integral domain and $R[t]^* = R^*$.
For this note that if $f, g \in R[t] \setminus \{0\}$ then $\deg(f), \deg(g) \geq 0$ and $\mathrm{lc}(f) \neq 0 \neq \mathrm{lc}(g)$. The degree formulae for polynomials in Proposition 6.16 then imply

$$\deg(f \cdot g) = \deg(f) + \deg(g) \geq 0, \tag{29}$$

since $\mathrm{lc}(f) \cdot \mathrm{lc}(g) \neq 0$ in the integral domain R. Hence, $f \cdot g \neq 0$ and $R[t]$ is an an integral domain. If $f \in R[t]^*$ is a unit and g the corresponding inverse then $f \cdot g = t^0 = 1$ and (29) implies that $\deg(f) = 0 = \deg(g)$. I.e. f and g are constant polynomials and therefore $f, g \in R^*$. If conversely $f \in R^* \subseteq R[t]$ then there is a $g \in R \subseteq R[t]$ with $f \cdot g = 1 = t^0$ and thus $f \in R[t]^*$.

f. $\mathbb{Z}_4$ is not an integral domain, since $\overline{2}$ is a zero divisor due to $\overline{2} \cdot \overline{2} = \overline{0}$.

Lemma 7.3 (Cancellation Rules)
If R is an integral domain then the cancellation rules for the multiplication hold, i.e. for all $a, b, c \in R$ with $a \neq 0$ we have

$$a \cdot b = a \cdot c \quad \Longrightarrow \quad b = c$$

and

$$b \cdot a = c \cdot a \quad \Longrightarrow \quad b = c.$$

Proof: Due to the commutativity of the multiplication it suffices to show one of the cancellation rules. Let $a, b, c \in R$ with $ab = ac$ be given. Then

$$0 = ab - ac = a \cdot (b - c). \tag{30}$$

Since $a \neq 0$ and R is an integral domain a cannot be a zero divisor. Thus (30) implies that $b - c = 0$ and hence $b = c$. □

We can now introduce the notion of divisibility for elements in an integral domain. Note here that on an integral domain we have the operations of addition and of multiplication but we do not have division! We thus must define divisibility with the help of the operation of multiplication. define.

Definition 7.4
Let R be an integral domain and $a, b \in R$.

a. We say b *divides* a if there is a $c \in R$ such that $a = b \cdot c$. We then write $b \mid a$.

b. We call $g \in R$ a *greatest common divisor* of a and b if the following two properties are fulfilled:
   (1) $g \mid a$ and $g \mid b$.
   (2) For all $h \in R$ with $h \mid a$ and $h \mid b$ we have $h \mid g$.
   We denote by

   $$\mathrm{GCD}(a, b) = \{g \in R \mid g \text{ is a greatest common divisor of } a \text{ and } b\}$$

   the set the greatest common divisors of a and b.

c. We call $k \in R$ a *lowest common multiple* of a and b if the following two properties are fulfilled:
   (1) $a \mid k$ and $b \mid k$.
   (2) For all $l \in R$ with $a \mid l$ and $b \mid l$ we have $k \mid l$.
   We denote by

   $$\mathrm{LCM}(a, b) = \{k \in R \mid k \text{ is a lowest common multiple of } a \text{ and } b\}$$

the set the lowest common multiples of a and b.

d. We call a and b *coprime* if $1 \in \mathrm{GCD}(a,b)$.

Remark 7.5

In the definition of a greatest common divisors g of a and b Condition (1) means that g is a divisor of a and of b at all. Condition (2) justifies the "*greatest*" in the definition. How can we possibly decide in an arbitrary integral domain R if g is larger than h for $g, h \in R$? In $\mathbb{Z}$ we can use the well known order relation "$\leq$", e.g. among the common divisors $h = 2$ and $g = 6$ of the integers $a = 12$ and $b = 30$ the larger one is 6. But how should we proceed in $\mathbb{Z}[t]$? Is $t + 2$ larger than t or smaller or can we simply not compare them? We therefore use another property of the integers to decide which of the common divisors is the largest and can thus be called the *greatest common divisor*. Namely, it will be divided by all other common divisors! In the above example 1, 2, 3 and 6 are the only positive common divisors of 12 and 30 and they all divide $g = 6$. We can thus call a divisor *smaller* than another one if it divides the other one. In this sense Condition (2) singles out the largest common divisor of g and h if $R = \mathbb{Z}$.

By definition a *greatest common divisor* of two elements a and b is a common divisor of a and b which is divisible by each common divisor. Analogously, a *lowest common multiple* of a and b is a common multiple which divides each other common multiple.

Why do we talk in the definition of *a* greatest common divisor and not of *the* greatest common divisor? Simply since our definition does not determine it uniquely! Again in the above example $a = 12, b = 30 \in \mathbb{Z}$ is $g = 6$ is obviously a greatest common divisor of a and b. However, also -6 divides a and b and it is divisible by every other common divisor of a and b. In the integers $\mathbb{Z}$ our definition determines a greatest common divisor only up to its sign (i.e. up to multiplication with a unit in $\mathbb{Z}$).

In $\mathbb{Q}[t]$ the situation becomes even worse. Consider two constant polynomials $0 \neq a, g \in \mathbb{Q} \subset \mathbb{Q}[t]$ then

$$a = g \cdot \frac{a}{g}$$

and therefore g is a divisor of a. Moreover, for each divisor $c \in \mathbb{Q}[t]$ of a there is a $d \in \mathbb{Q}[t]$ such that $a = c \cdot d$, and from the degree formula $0 = \deg(a) = \deg(c) + \deg(d)$ we then deduce that $\deg(c) = 0$ and $c \in \mathbb{Q} \setminus \{0\}$. I.e. the divisors of a are exactly the elements in $\mathbb{Q} \setminus \{0\}$. If we consider in $\mathbb{Q}[t]$ the constant polynomials $a = 2$ and $b = 5$ then the rational numbers $0 \neq q \in \mathbb{Q}$ are exactly the common divisors of a and b and since they all divide each other they will all be common divisors of a and b in the sense of our definition. Since one can move from one common divisor q to another one p by multiplication with the rational number $\frac{p}{q}$ we can say that the greatest common divisor is only determined up to multiplication by a non-zero rational number.

In the integers one commonly prefers the positive greatest common divisor, but for an arbitrary integral domain there is no obvious choice among the greatest common divisors and we therefore consider the set $\mathrm{GCD}(a, b)$ of all greatest common divisors.

The considerations for the greatest common divisor apply analogously to a a lowest common multiple k of a and b. Condition (1) means that k is indeed a multiple of both a and b, and (2) justifies the "*lowest*" since each other multiple is divisible by the lowest one. In $\mathbb{Z}$ the lowest common multiple is only determined up to its sign. □

Example 7.6

a. For $f = t - 1 \in \mathbb{Q}[t]$ and $g = t^n - 1 \in \mathbb{Q}[t]$ with $n \geq 1$ we have

$$g = f \cdot \left(t^{n-1} + t^{n-2} + \ldots + t + 1\right)$$

and thus $f \mid g$.

b. Consider the complex numbers $a = 9$, $b = 2 + \sqrt{-5}$, $c = 2 - \sqrt{-5}$ and $d = 3$ in $\mathbb{Z}\left[\sqrt{-5}\right]$. Due to

$$a = 9 = \left(2 + \sqrt{-5}\right) \cdot \left(2 - \sqrt{-5}\right) = b \cdot c$$

we have $b \mid a$.

We now want to show that d is not a divisor of b. Suppose the contrary, i.e. $d \mid b$. Then there is an $e = x + y \cdot \sqrt{-5}$ with $x, y \in \mathbb{Z}$ such that

$$b = d \cdot e.$$

With the aid of the absolute value of complex numbers we get

$$9 = |b|^2 = |d|^2 \cdot |e|^2 = 9 \cdot \left(x^2 + 5 \cdot y^2\right).$$

It follows that $x^2 + 5y^2 = 1$, and since x and y are integers we must have $y = 0$ and $x \in \{1, -1\}$. But then $b \notin \{d, -d\}$ so that we have deduced a contradiction.

c. In $\mathbb{Z}$ we have $\mathrm{GCD}(6, 8) = \{-2, 2\}$ and $\mathrm{LCM}(6, 8) = \{-24, 24\}$.

Many properties of an element element a in an integral domain can be expressed by properties of the ideal generated by a, i.e. by $\langle a \rangle_R = \{r \cdot a \mid r \in R\}$. Sometimes arguments become shorter that way.

Lemma 7.7

Let R be an integral domain and $a, b, g, k \in R$.

a. $b \mid a$ *if and only if* $\langle a \rangle_R \subseteq \langle b \rangle_R$.

b. *The following statements are equivalent:*

(i) $a \mid b$ *and* $b \mid a$.

(ii) $\langle a \rangle_R = \langle b \rangle_R$.

(iii) *There is a unit* $u \in R^*$ *with* $a = u \cdot b$.

c. $g \in \mathrm{GCD}(a, b)$ *if and only if the following two properties are fulfilled:*

(1) $\langle a, b\rangle_R \subseteq \langle g\rangle_R$.

(2) *For all* $h \in R$ *with* $\langle a, b\rangle_R \subseteq \langle h\rangle_R$*, we have* $\langle g\rangle_R \subseteq \langle h\rangle_R$.

d. $k \in \mathrm{LCM}(a, b)$ *if and only if the following two properties are fulfilled:*

(1) $\langle k\rangle_R \subseteq \langle a\rangle_R \cap \langle b\rangle_R$.

(2) *For all* $l \in R$ *with* $\langle l\rangle_R \subseteq \langle a\rangle_R \cap \langle b\rangle_R$*, we have* $\langle l\rangle_R \subseteq \langle k\rangle_R$.

Proof: a. If $b \mid a$ then there is a $c \in R$ with $a = b \cdot c$. Hence we have for each $r \in R$ also $r \cdot a = (r \cdot c) \cdot b \in \langle b\rangle_R$ and therefore $\langle a\rangle_R \subseteq \langle b\rangle_R$. Conversely if we have $\langle a\rangle_R \subseteq \langle b\rangle_R$ then $a \in \langle a\rangle_R \subseteq \langle b\rangle_R$ and hence there is a $c \in R$ with $a = c \cdot b$. Therefore b divides a.

b. First we may assume without loss of generality that $a \neq 0 \neq b$ since the three statements are obviously satisfied otherwise. Suppose first that (i) holds then (ii) follows from a.. If (ii) holds then $a \in \langle b\rangle_R$ and $b \in \langle a\rangle_R$. Therefore there are $u, v \in R$ with $a = u \cdot b$ and $b = v \cdot a$. But then we have

$$1 \cdot a = a = u \cdot b = (u \cdot v) \cdot a.$$

Since in the integral domain R the cancellation rules hold and since $a \neq 0$ we get $1 = u \cdot v$. Since moreover R is commutative $u \in R^*$ is a unit and by the choice of u we have $a = u \cdot b$ so that (iii) is fulfilled. If we now suppose that (iii) holds then we have $a = u \cdot b$ and $b = u^{-1} \cdot a$. This implies $b \mid a$ and $a \mid b$ so that (i) is fulfilled.

c. This is just a reformulation of the definition with the aid of Part a. — note for this that $\langle a, b\rangle_R \subseteq \langle g\rangle_R$ if and only if $\langle a\rangle_R \subseteq \langle g\rangle_R$ and $\langle b\rangle_R \subseteq \langle g\rangle_R$.

d. This is just a reformulation of the definition with aid of Part a..

□

A generalisation of the considerations on the GCD and the LCM in $\mathbb{Z}$ and $\mathbb{Q}[t]$ is the following Lemma. Two greatest common divisors differ only by a unit, and the same holds true for two lowest common multiples.

Lemma 7.8

Let R *be an integral domain,* $a, b \in R$.

a. *If* $g \in \mathrm{GCD}(a, b)$ *then* $\mathrm{GCD}(a, b) = \{u \cdot g \mid u \in R^*\}$*, i.e. a greatest common divisor is only up to multiplication with units determined.*

b. *If* $k \in \mathrm{LCM}(a, b)$*, then* $\mathrm{LCM}(a, b) = \{u \cdot k \mid u \in R^*\}$*, i.e. a lowest common multiple is only up to multiplication with units determined.*

Proof: The proof is left as an exercise for the reader. □

Exercise 7.9

Prove Lemma 7.8.

Exercise 7.10

a. Find the zero divisors and the units in $\mathbb{Z}_{24}$. Is $\mathbb{Z}_{24}$ an integral domain?

b. Is $3+4i$ a divisor of $7+i$ in $\mathbb{Z}[i]$?

c. Find all greatest common divisors of $f = t^2 - 3t + 2$ and $g = t^3 - 2t^2 - t + 2$ in $\mathbb{Z}[t]$.

The next Exercise shows that the integers the greatest common divisor respectively the lowest common multiple can be defined using the order relation on $\mathbb{Z}$.

Exercise 7.11
Let $a, b \in \mathbb{Z}$ be two integers. In Notation 4.43 we have introduced the number

$$\operatorname{lcm}(a,b) := \begin{cases} \min\{z > 0 \mid a \text{ and } b \text{ divide } z\}, & \text{if } a, b \neq 0, \\ 0, & \text{if } a = 0 \text{ or } b = 0, \end{cases}$$

and we now complement it by

$$\gcd(a,b) := \begin{cases} \max\left\{z > 0 \;\middle|\; z \text{ divides } a \text{ as well as } b\right\}, & \text{if } (a,b) \neq (0,0), \\ 0, & \text{otherwise.} \end{cases}$$

Show that $\gcd(a,b) \in \mathrm{GCD}(a,b)$ and $\operatorname{lcm}(a,b) \in \mathrm{LCM}(a,b)$.

Exercise 7.12
Let R be a commutative ring with one which contains only a finite number of elements. Show that each element of R is either a unit or a zero divisor.

Exercise 7.13
Let R be an integral domain. We define an equivalence relation on $R \times (R \setminus \{0\})$ by

$$(a,b) \sim (a',b') \quad :\Longleftrightarrow \quad a \cdot b' = a' \cdot b.$$

The equivalence class of (a,b) is denoted by $\frac{a}{b}$, and the set of all equivalence classes is denoted by $\mathrm{Quot}(R)$. On this set we define an addition and a multiplication by

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

Show:

a. $\sim$ is an equivalence relation.

b. The addition and the multiplication are well defined.

c. $\big(\mathrm{Quot}(R), +, \cdot\big)$ is a field, the so called *field of fractions* of R.

B) **Factorial rings**

Up to now we have studied neither the question whether a greatest common divisor of two elements in an integral domain always exists nor whether we are able to compute it if it exists. We will indeed see later that there are integral domains where greatest common divisors do not necessarily exist. But we want to concentrate first on the question of their computation.

In school the most common method to compute a greatest common divisor of two integers would be to decompose the integers as a product of prime numbers and compare which prime factors with what multiplicity they share. If we want to copy this algorithm for other rings then we need to generalise the notion of a *prime number*. For this we have to study characterising properties of prime numbers. One commonly defines a *prime number* to be *positive* integer, which has precisely two positive divisors. This can be expressed more formally as $p \in \mathbb{Z}_{>1}$ is a prime number if and only if for $a, b \in \mathbb{Z}_{\geq 0}$ we have:

$$p = a \cdot b \quad \Longrightarrow \quad a = 1 \;\text{ or }\; b = 1. \tag{31}$$

For this note that $p = a \cdot b$ means that a and b are both divisors of p and for a prime number not both can be p.

There is another property which characterises prime numbers. It is a property that is used when we compute the greatest common divisor in the above mentioned manner. Namely, if a prime number divides a product then it divides already one of its factors. I.e. $p \in \mathbb{Z}_{>1}$ is a prime number if and only if for $a, b \in \mathbb{Z}_{\geq 0}$ we have:

$$p \mid a \cdot b \quad \Longrightarrow \quad p \mid a \;\text{ or }\; p \mid b. \tag{32}$$

The proof of the equivalence of the properties (31) and (32) is given in Corollary 7.18.

If we now want to generalise the notion of prime number to an arbitrary integral domain we have two possibilities to do so. We will see that these two notions do not necessarily coincide. However, only if they coincide we can expect to get a sensible theory of divisibility since only then we will be able to generalise the unique prime factorisation that we know from the integers.

One problem in the generalisation of the above conditions to an arbitrary integral domain seems to be the lack of the order relation $>$, which would allow us to talk about *positive* ring elements. This however turns out to be superfluous if we simply allow in $\mathbb{Z}$ also negative divisors. The condition "$= 1$" respectively "$\neq 1$" can then be replaced by "$\in \mathbb{Z}^*$" respectively "$\notin \mathbb{Z}^*$".

Definition 7.14
Let R be an integral domain.

a. An element $0 \neq p \in R \setminus R^*$ is called *irreducible* if $p = a \cdot b$ with $a, b \in R$ implies that $a \in R^*$ or $b \in R^*$.

b. An element $0 \neq p \in R \setminus R^*$ is called *prime* if $p \mid a \cdot b$ with $a, b \in R$ implies that $p \mid a$ or $p \mid b$.

c. R is called a *factorial* or a *unique factorisation domains* if each $0 \neq a \in R \setminus R^*$ can be decomposed as a product of finitely man prime elements.

We will see later that in a factorial ring the decomposition into a product of prime elements is essentially unique.

Example 7.15

a. We distinguish in $\mathbb{Z}$ between *prime numbers* (which are *positive* by definition) and *prime elements* (which are allowed to be negative). Due to the above considerations an integer z is prime if and only if it is irreducible, and this is the case if and only if z or $-z$ is a prime number. As indicated we will prove this fact Corollary 7.18.

b. If K is a field and $f \in K[t]$ with $\deg(f) = 1$ then f is irreducible.

   To see this consider the decomposition $f = g \cdot h$ with $g, h \in K[t]$. The degree formulae then give $1 = \deg(f) = \deg(g) + \deg(h)$. Hence we have either $\deg(g) = 0$ and $\deg(h) = 1$ or we have $\deg(g) = 1$ and $\deg(h) = 0$. In the first case $g \in K \setminus \{0\} = K^* = K[t]^*$ while in the latter case $h \in K \setminus \{0\} = K^* = K[t]^*$, where the equality $K \setminus \{0\} = K^*$ comes from the fact that K is a field.

c. The polynomial $f = 2t + 2 \in \mathbb{Z}[t]$ is not irreducible since $f = 2 \cdot (t + 1)$ and neither 2 nor $t + 1$ is a unit in $\mathbb{Z}[t]$ since $\mathbb{Z}[t]^* = \mathbb{Z}^* = \{1, -1\}$.

d. If R is an integral domain and are $p, q \in R$ are irreducible with $p \mid q$ then $\langle p \rangle_R = \langle q \rangle_R$, i.e. the two differ only by a unit.

   To see this note that $p \mid q$ means that there exists a $c \in R$ with $q = p \cdot c$. Since q is irreducible and p is not a unit necessarily c must be a unit. Therefore p and q differ only by a unit.

We now want to study the connection between the notions *prime* and *irreducible* in more detail, and we want to show that the two notions coincide in the ring of integers.

Lemma 7.16
If R is an integral domain and $p \in R$ prime then p is irreducible.

Proof: Let $a, b \in R$ given with $p = a \cdot b$ then we have In particular $p \mid a \cdot b$. Hence we have $p \mid a$ or $p \mid b$. In the first case there is a $c \in R$ with $a = p \cdot c$ and thus we have

$$p \cdot 1 = p = a \cdot b = p \cdot c \cdot b.$$

Since in the integral domain R the cancellation rules hold we get that $1 = c \cdot b$ and b is a unit. Analogously from $p \mid b$ it follows that $a \in R^*$. Thus p is irreducible. $\square$

Example 7.17

a. We now want to give an example that an irreducible element need not be prime.

   Consider again the complex numbers $a = 9$, $b = 2 + \sqrt{-5}$, $c = 2 - \sqrt{-5}$ and $d = 3$ in $\mathbb{Z}\big[\sqrt{-5}\big]$. We have already seen in Example 7.6 that d is not a divisor of b. Analogously one shows that d is not a divisor of c. But $d = 3$ is a divisor of $d^2 = a = b \cdot c$. Hence d *not prime* since it divides the product $b \cdot c$ but none of its factors.

Let now $d = f \cdot g$ with $f = x + y \cdot \sqrt{-5}$ and $g = u + v \cdot \sqrt{-5}$, $x, y, u, v \in \mathbb{Z}$. Then we have

$$9 = |d|^2 = |f|^2 \cdot |g|^2 = \left(x^2 + 5y^2\right) \cdot \left(u^2 + 5v^2\right)$$

with $x^2+5y^2, u^2+5v^2 \in \mathbb{N}$. We deduce that $(x^2+5y^2, u^2+5v^2) \in \{(9,1),(1,9)\}$. In the first case we get necessarily $u \in \{1,-1\}$ and $v = 0$ so that g is a unit in $\mathbb{Z}\left[\sqrt{-5}\right]$. In the second case we have $x \in \{1,-1\}$ and $y = 0$ so that f is a unit in $\mathbb{Z}\left[\sqrt{-5}\right]$. Hence, d is *irreducible*.

b. If R is factorial then each irreducible element is prime.

For this note that if $p \in R$ is irreducible then p by assumption p is a product $p = q_1 \cdots q_k$ of prime elements. Suppose that $k \geq 2$. Since q_k is prime it is not a unit and hence necessarily $q_1 \cdots q_{k-1}$ is a unit. Therefore there is an $a \in R$ with $1 = q_1 \cdot (q_2 \cdots q_{k-1} \cdot a)$ so that also q_1 is a unit in contradiction to the assumption that q_1 is prime. Hence $k = 1$ and $p = q_1$ is prime.

c. $\mathbb{Z}\left[\sqrt{-5}\right]$ is not factorial since 3 is irreducible but not prime.

So far we have know only a single field with a finite number of elements, $\mathbb{Z}_2$. For applications in cryptography and coding theory, however, finite fields are rather important. We will therefore show that the example $\mathbb{Z}_2$ can be generalised.

Corollary 7.18
For $0 \neq n \in \mathbb{Z}$ the following statements are equivalent:

a. $\mathbb{Z}_n$ *is a field.*

b. $\mathbb{Z}_n$ *is an integral domain.*

c. n *is prime.*

d. n *is irreducible, i.e. n is a prime number.*

Proof:

a. $\Rightarrow$ b.: If $\mathbb{Z}_n$ is a field then $\mathbb{Z}_n$ an integral domain by Example 7.2.

b. $\Rightarrow$ c.: If we have $n \mid a \cdot b$ with $a, b \in \mathbb{Z}$ then

$$\overline{a} \cdot \overline{b} = \overline{a \cdot b} = \overline{0} \in \mathbb{Z}_n.$$

Since by assumption $\mathbb{Z}_n$ is an integral domain we must have $\overline{a} = \overline{0}$ or $\overline{b} = \overline{0}$. In the first case we have $n \mid a$, in latter $n \mid b$. Therefore n is prime in $\mathbb{Z}$.

c. $\Rightarrow$ d.: This follows from Lemma 7.16.

d. $\Rightarrow$ a.: If I is an ideal in $\mathbb{Z}_n$ then $(I, +)$ a subgroup of $(\mathbb{Z}_n, +)$ and the order of I is by the Theorem of Lagrange a divisor the prime number $n = |\mathbb{Z}_n|$. Therefore $|I|$ must either be 1 or n, i.e. $I = \{\overline{0}\}$ or $I = \mathbb{Z}_n$. Since moreover $\{\overline{0}\} \neq \mathbb{Z}_n$ the commutative ring $\mathbb{Z}_n$ with one has therefore exactly two ideals and is by Exercise 6.36 a field.

□

Remark 7.19
If R is a factorial ring and $0 \neq a \in R \setminus R^*$ then the representation $a = p_1 \cdots p_r$ as a product of prime elements is *essentially unique*, i.e. if

$$p_1 \cdot \cdots \cdot p_r = q_1 \cdot \cdots \cdot q_s \tag{33}$$

are two such representations we have $r = s$ and after possibly renumbering the q_i the elements p_i and q_i differ only by a unit, i.e. $\langle p_i \rangle_R = \langle q_i \rangle_R$. This can easily be seen: since p_1 prime and since it is a divisor of the right hand side of (33) p_1 must divide one of the q_i. Renumbering the q_i we can assume that $p_1 \mid q_1$. Since both are prime the are also irreducible by Lemma 7.16 and therefore differ only by a unit by Example 7.15. We thus can cancel p_1 in both sides of (33) and proceed inductively. □

Remark 7.20
Let R be a factorial ring and $a = p_1^{m_1} \cdots p_r^{m_r}$ and $b = p_1^{n_1} \cdots p_r^{n_r}$ be elements in $R \setminus \{0\}$ where $p_1, \ldots, p_r$ are prime, $\langle p_i \rangle_R \neq \langle p_j \rangle_R$ for $i \neq j$ and $m_1, \ldots, m_r, n_1, \ldots, n_r \in \mathbb{N}$. Then one gets as in the case of the ring of integers that

$$p_1^{\min\{m_1,n_1\}} \cdots p_r^{\min\{m_r,n_r\}} \in \mathrm{GCD}(a, b)$$

and

$$p_1^{\max\{m_1,n_1\}} \cdots p_r^{\max\{m_r,n_r\}} \in \mathrm{LCM}(a, b).$$

We call a representation of a as above as above also a *prime factorisation* of a if $m_i > 0$ for all $i = 1, \ldots, r$. By Remark 7.19 it is up to the order of the factors and up to multiplication with units uniquely determined. □

Exercise 7.21
Let R be an integral domain and suppose there is a natural number $n \geq 1$ such that $n \cdot 1_R = \sum_{k=1}^{n} 1_R = 0_R$, i.e. the n-fold sum of the 1-element is zero. Show that the smallest positive integer $p = \min\{m \in \mathbb{Z}_{>0} \mid m \cdot 1_R = 0_R\}$ with this property is irreducible (i.e. a prime number).

We call this number p the *characteristic* of the ring.

Exercise 7.22
If $p = x + y \cdot i \in \mathbb{Z}[i]$ with $q := |p|^2 = x^2 + y^2$ being a prime number then p is a prime element in $\mathbb{Z}[i]$. Find an example for a such number p.

Exercise 7.23
Find all polynomials f in $\mathbb{Z}_2[t]$ of degree 4 whose leading coefficient $\mathrm{lc}(f)$ and whose constant coefficient $f(0)$ both are $\overline{1}$. Which of these polynomials are irreducible?

Exercise 7.24

a. Let $f \in \mathbb{Z}[t]$ with $\mathrm{lc}(f) = 1$. Show if there is a prime number $p \in \mathbb{Z}$ such that the reduction $\phi_p(f)$ of f modulo p irreducible in $\mathbb{Z}_p[t]$ (see Exercise 6.41) then f is irreducible in $\mathbb{Z}[t]$.

b. Find all polynomials f in $\mathbb{Z}_2[t]$ of degree $0 \leq \deg(f) \leq 4$ and write them as products of as many polynomials of of degree at least one as possible.

c. Is $f = t^4 + 187t^3 + 5t^2 - 33t + 3001 \in \mathbb{Z}[t]$ irreducible?

C) **Euclidean Rings**

Factorial rings generalise the integers and as we have seen in Remark 7.20 the *unique prime factorisation* of an element is very useful. However, so far we do not know any other factorial ring besides the ring of integers. We do not yet know of any good criterion to decide positively that a certain ring is factorial.

Let us postpone this problem for a moment. We started the discussion on factorial rings with our wish for an algorithm to be able to compute a greatest common divisor. This can be achieved by computing prime factorisations. However, even in the ring of integers we do not have an efficient algorithm for the latter. You certainly would not want to try it by hand on the following integers:

$$a = 1234567890987654321234567890987654321$$

and

$$b = 2728395039082716049928395039082706 5.$$

Even for your pocket calculator this would be more than only a challenge. But for applications in cryptography these two integers are indeed awfully small! There integers with 500 and more digits are needed to create secure cryptographical procedures. Their security very much depends on the fact that it is difficult to factorise these large integers.

However, there is a very simple and efficient algorithm for computing a greatest common divisor in the integers, and it does not need at all a prime factorisation. We will describe this algorithm in the sequel. It works in each integral domain where we can do a kind of *division with remainder*.

Definition 7.25
An integral domain R is called a *Euclidean ring*, if there is a function

$$\nu : R \setminus \{0\} \longrightarrow \mathbb{N},$$

such that for all $a, b \in R \setminus \{0\}$ there is a *division with remainder* the form

$$a = q \cdot b + r$$

with $q, r \in R$ where either $r = 0$ or $\nu(r) < \nu(b)$. We call ν a *Euclidean function* of R.

Example 7.26
$\mathbb{Z}$ is a Euclidean ring with Euclidean function $\nu : \mathbb{Z} \setminus \{0\} \longrightarrow \mathbb{N} : z \mapsto |z|$ due to the well known division with remainder (see Remark 1.37).

Before we show how the division with remainder helps to compute a greatest common divisor we want to show that there are other Euclidean rings besides the ring of integers. The most important class of examples are the polynomial rings over fields.

Proposition 7.27 (Division with Remainder in the Polynomial Ring)
If K is a field and $0 \neq f, g \in K[t]$ then there are polynomials $q, r \in K[t]$ such that

$$g = q \cdot f + r \quad \text{and} \quad \deg(r) < \deg(f).$$

The polynomials q and r are uniquely determined.

Proof: Let $f = \sum_{i=0}^{n} a_i \cdot t^i$ and $g = \sum_{i=0}^{m} b_i \cdot t^i$ with $m = \deg(g)$ and $n = \deg(f)$. We do the proof of the existence of such a division with remainder by induction by m.

If $m = n = 0$ we are done by $q = \frac{b_0}{a_0}$ and $r = 0$, and if $0 \leq m < n$ we can choose $q = 0$ and $r = g$. These cases include the induction basis $m = 0$.

It suffices to consider the case $m > 0$ and $n \leq m$. We define

$$g' := g - \frac{b_m}{a_n} \cdot t^{m-n} \cdot f.$$

Then the leading coefficients of the two summands cancel out so that $\deg(g') < \deg(g) = m$. Therefore, by induction there exist polynomials $q', r' \in K[t]$ such that

$$q' \cdot f + r' = g' = g - \frac{b_m}{a_n} \cdot t^{m-n} \cdot f$$

and $\deg(r') < \deg(f)$. Therefore

$$g = \left(q' + \frac{b_m}{a_n} \cdot t^{m-n}\right) \cdot f + r',$$

and we are done with $q = q' + \frac{b_m}{a_n} \cdot t^{m-n}$ and $r = r'$.

It remains to show the uniqueness of the decomposition. Suppose we have two such decompositions

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

with $q, q', r, r' \in K[t]$ and $\deg(r), \deg(r') < \deg(f)$. Then we have

$$(q - q') \cdot f = r' - r$$

and hence

$$\deg(q - q') + \deg(f) = \deg(r' - r) \leq \max\{\deg(r), \deg(r')\} < \deg(f).$$

The is only possible if $q - q' = 0$. Therefore we have $q = q'$ and then also $r = r'$. $\square$

We get the following corollary as an immediate consequence.

Corollary 7.28
If K a field then $K[t]$ is a Euclidean ring with deg *as Euclidean function.*

The proof of Proposition 7.27 is constructive, i.e. it gives us an algorithm for the computation of the division with remainder in the polynomial ring.

Example 7.29
Let $f = t^3 + t + 1, g = t - 1 \in \mathbb{Q}[t]$ be given. We do polynomial division

$$\begin{array}{l}
(t^3 \qquad + \ t + 1) : (t-1) = t^2 + t + 2 + \frac{r}{t-1} \\
\underline{t^3 - t^2} \\
\qquad t^2 + \ t \\
\qquad \underline{t^2 - \ t} \\
\qquad\qquad 2t + \ 1 \\
\qquad\qquad \underline{2t - \ 2} \\
\qquad\qquad\quad 3 \quad =: r
\end{array}$$

and get $f = (t^2 + t + 2) \cdot g + 3$. □

We now want to get to know the Euclidean Algorithm which allows us to compute a greatest common divisor in a Euclidean ring. Before we formulate the algorithm as a theorem we want to demonstrate it at an example.

Example 7.30
We want to compute a greatest common divisor of the integers $r_0 = 66$ and $r_1 = 15$. For this we do division with remainder by

$$r_0 = 66 = 4 \cdot 15 + 6 = q_1 \cdot r_1 + r_2$$

and get the remainder $r_2 = 6$. Then we divide r_1 by r_2 with remainder,

$$r_1 = 15 = 2 \cdot 6 + 3 = q_2 \cdot r_2 + r_3,$$

and get the remainder $r_3 = 3$. Then divide r_2 by r_3 with remainder,

$$r_2 = 6 = 2 \cdot 3 + 0 = q_2 \cdot r_3 + r_4,$$

and get the remainder $r_4 = 0$. The algorithm terminates since r_3 cannot be divided by $r_4 = 0$. We claim that then

$$r_3 = 3 \in \mathrm{GCD}(66, 15) = \mathrm{GCD}(r_0, r_1),$$

is greatest common divisor of $r_0 = 66$ and $r_1 = 15$, i.e. the last non-zero remainder in the successive division with remainder is a greatest common divisor.

The following theorem explains why the above algorithm works.

Theorem 7.31 (Euclidean Algorithm)
Let R be a Euclidean ring with Euclidean function ν and $r_0, r_1 \in R \setminus \{0\}$.

For $k \geq 2$ and as long as $r_{k-1} \neq 0$ define recursively r_k as a remainder of the division with remainder of r_{k-2} by r_{k-1}, i.e. there is a $q_{k-1} \in R$ with

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

with

$$r_k = 0 \quad \text{or} \quad \nu(r_k) < \nu(r_{k-1}).$$

Then there is an $n \geq 2$ such that $r_n = 0$, and we have $r_{n-1} \in \mathrm{GCD}(r_0, r_1)$.

Proof: As long as r_{k-1} is not zero we can go on with the division with remainder and get that way integers $r_k, q_{k-1} \in R$ such that the above conditions are fulfilled .

Our construction gives

$$\begin{aligned} r_0 &= r_1 q_1 + r_2, & \nu(r_2) < \nu(r_1), \\ r_1 &= r_2 q_2 + r_3, & \nu(r_3) < \nu(r_2), \\ &\vdots \\ r_{k-2} &= r_{k-1} q_{k-1} + r_k, & \nu(r_k) < \nu(r_{k-1}), \end{aligned}$$

and therefore we get a strictly decreasing sequence of natural numbers

$$\nu(r_1) > \nu(r_2) > \nu(r_3) > \ldots .$$

This, however, shows that the algorithm must stop, i.e. there is an $n \geq 2$ such that $r_n = 0$.

We now want to show by induction on n that

$$r_{n-1} \in \mathrm{GCD}(r_0, r_1).$$

Note that $n - 1$ is the number of recursion steps of the algorithm, i.e. the number of divisions with remainder which are computed until the remainder zero occurs.

Induction basis: $n = 2$. Then $r_2 = 0$ and therefore $r_1 \mid r_0$ and $r_1 \in \mathrm{GCD}(r_0, r_1)$.

Induction step: Let $n \geq 3$ and suppose the statement holds for all pairs for which the algorithm finishes with at least one step less. If we only consider the last $n - 3$ recursion steps the induction hypothesis applied to r_1 and r_2 shows:

$$r_{n-1} \in \mathrm{GCD}(r_1, r_2).$$

In particular r_{n-1} is a divisor of r_1 and of r_2. Since by assumption $r_0 = q_1 \cdot r_1 + r_2$ then r_{n-1} is also a divisor of r_0.

Let now $r \in R$ be a further divisor of r_0 and r_1. Then we have

$$r \mid (r_0 - q_1 \cdot r_1) = r_2,$$

and hence r is a divisor of both r_1 and r_2. But since $r_{n-1} \in \mathrm{GCD}(r_1, r_2)$ we have

$$r \mid r_{n-1},$$

and by definition we get $r_{n-1} \in \mathrm{GCD}(r_0, r_1)$. □

In order to apply the algorithm we only need division with remainder. Since we have such a division for each polynomial ring over a field we can compute greatest common divisors there.

Example 7.32

Consider $r_0 = t^4 + t^2 \in \mathbb{Q}[t]$ and $r_1 = t^3 - 3t^2 + t - 3 \in \mathbb{Q}[t]$. Division with remainder give in the first step

$$r_0 = t^4 + t^2 = (t + 3) \cdot (t^3 - 3t^2 + t - 3) + (9t^2 + 9) = q_1 \cdot r_1 + r_2$$

with remainder $r_2 = 9t^2 + 9$. In the second step we get

$$r_1 = t^3 - 3t^2 + t - 3 = \left(\frac{1}{9} \cdot t - \frac{1}{3}\right) \cdot (9t^2 + 9) + 0 = q_2 \cdot r_2 + r_3$$

with remainder $r_3 = 0$. The algorithm stops and $r_2 = 9t^2 + 9 \in \mathrm{GCD}(r_0, r_1)$ is a greatest common divisor. We can normalise this polynomial by dividing it by its leading coefficient. That way we get the monic greatest common divisor

$$t^2 + 1 \in \mathrm{GCD}\left(t^4 + t^2, t^3 - 3t^2 + t - 3\right).$$

Remark 7.33
If R is a Euclidean ring and $a, b \in R \setminus \{0\}$ we have $g \in \mathrm{GCD}(a, b)$ if and only if $\frac{a \cdot b}{g} \in \mathrm{LCM}(a, b)$. Thus we can also compute lowest common multiples with the aid of the Euclidean Algorithm.

Exercise 7.34
Show that $\mathbb{Z}[i] = \{x + i \cdot y \mid x, y \in \mathbb{Z}\}$ is a Euclidean ring with Euclidean function $\nu : \mathbb{Z}[i] \longrightarrow \mathbb{N} : a \mapsto |a|^2$.

Exercise 7.35
Consider the polynomials

$$f = t^5 + 3t^4 + 2t^3 + 5t^2 + 7t + 2 \in \mathbb{Z}[t]$$

and

$$g = t^3 + t^2 + t + 1 \in \mathbb{Z}[t].$$

a. Find a greatest common divisor of f and g in $\mathbb{Q}[t]$ using the Euclidean Algorithm.
b. Consider the coefficients of f and g modulo 3 and find a greatest common divisor of the resulting polynomials $\phi_3(f)$ and $\phi_3(g)$ in $\mathbb{Z}_3[t]$.

D) The Polynomial Ring $K[t]$

In the last subsection we have seen how polynomial division works and that it actually is a division with remainder which gives the polynomial ring $K[t]$ over a field K the structure of a Euclidean ring. In this subsection we want to apply the division with remainder in order to factorise a polynomial which has a zero.

Lemma 7.36 (Substitution Morphism)
Let K a field and $b \in K$. The map

$$\varphi_b : K[t] \longrightarrow K : f \mapsto f(b)$$

is a ring epimorphism where

$$f(b) := \sum_{k=0}^{n} a_k \cdot b^k \in K$$

for $f = \sum_{k=0}^{n} a_k \cdot t^k \in K[t]$. We call φ_b substitution morphism, *and for a constant polynomial $f = a_0$ we have $\varphi_b(f) = a_0$.*

Proof: See Exercise 6.26. □

Remark 7.37
The fact that φ_b is a ring homomorphism implies

$$(f+g)(b) = f(b) + g(b) \quad \text{and} \quad (f \cdot g)(b) = f(b) \cdot g(b).$$

Now we can define the notion of a zero of a polynomial.

Definition 7.38
Let K be a field and $b \in K$. b is called *zero* of f if $f(b) = \varphi_b(f) = 0$.

Proposition 7.39
Let K be a field and let $b \in K$ be a zero of the polynomial $0 \neq g \in K[t]$ then there is a $q \in K[t]$ with

$$g = q \cdot (t - b).$$

We call $t - b$ a linear factor *of the polynomial g.*

Proof: Division with remainder 7.27 applied to g and $f = t - b$ gives two polynomials $q, r \in K[t]$ with

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

and $\deg(r) < \deg(f) = 1$. The degree condition implies that that $r = r_0 \cdot t^0$ is a constant polynomial. Since $\varphi_b(f) = b - b = 0$ and since b is a zero of g we have

$$r_0 = \varphi_b(r) = \varphi_b(g - q \cdot f) = \varphi_b(g) - \varphi_b(q) \cdot \varphi_b(f) = 0.$$

Therefore, r is the zero polynomial and $g = q \cdot (t - b)$. □

Corollary 7.40
If K is a field and $0 \neq f \in K[t]$ is a polynomial of degree $\deg(f) \geq 2$ and if f has a zero in K then f is not irreducible.

Proof: If $b \in K$ is a zero of f then due to Proposition 7.39 there is a $q \in K[t]$ with $f = q \cdot (t - b)$. The degree formulae show

$$\deg(q) = \deg(f) - 1 \geq 1,$$

so that the two factors q and $t - b$ cannot be units in $K[t]$. Thus f is not irreducible. □

Example 7.41

a. Let $f = t^3 + t^2 - 5t - 2 \in \mathbb{Q}[t]$ then $f(2) = 0$ and polynomial division gives:

$$\begin{array}{l}
(t^3 + \ \ t^2 - 5t - 2) : (t-2) = t^2 + 3t + 1. \\
\underline{\ t^3 - 2t^2\ } \\
\qquad\ \ 3t^2 - 5t \\
\qquad\ \ \underline{3t^2 - 6t} \\
\qquad\qquad\quad\ t - \ 2 \\
\qquad\qquad\quad\ \underline{t - \ 2} \\
\qquad\qquad\qquad\ \ -
\end{array}$$

Therefore we have $f = (t^2 + 3t + 1) \cdot (t - 2)$ and f is not irreducible.

b. If $f = t^5 + t^4 + t^3 + t^2 + t + \overline{1} \in \mathbb{Z}_2[t]$ then we have

$$f\big(\overline{1}\big) = \overline{1} + \overline{1} + \overline{1} + \overline{1} + \overline{1} + \overline{1} = \overline{6} = \overline{0} \in \mathbb{Z}_2.$$

Therefore $\overline{1}$ is a zero of f and f is not irreducible. We can split of the linear factor $t - \overline{1}$ by polynomial division. Note that $\overline{1} = -\overline{1}$ in $\mathbb{Z}_2$, therefore $t - \overline{1} = t + \overline{1}$:

$$\begin{array}{l}
\big(t^5 + t^4 + t^3 + t^2 + t + \overline{1}\big) : \big(t + \overline{1}\big) = t^4 + t^2 + \overline{1} \\
\underline{\;t^5 + t^4 \qquad\qquad\qquad\qquad\quad} \\
\qquad\qquad t^3 + t^2 + t + \overline{1} \\
\qquad\qquad \underline{t^3 + t^2 \qquad\quad\;} \\
\qquad\qquad\qquad\qquad t + \overline{1} \\
\qquad\qquad\qquad\qquad \underline{t + \overline{1}} \\
\qquad\qquad\qquad\qquad\quad\; -
\end{array}$$

Therefore, we have $f = \big(t^4 + t^2 + \overline{1}\big) \cdot \big(t + \overline{1}\big)$.

Theorem 7.42
If K is a field and $0 \neq f \in K[t]$ is a polynomial then f has at most $\deg(f)$ zeros.

Proof: We leave the prof as an exercise for the reader. □

Example 7.43
The polynomial $f = t^2 + 1$ has no zero in $\mathbb{R}$ while it has the zeros i and $-i$ in $\mathbb{C}$. Hence in $\mathbb{C}[t]$ it factors into product of linear factors:

$$f = (t - i) \cdot (t + i).$$

Remark 7.44
By the definition the substitution morphism defines for defines for each polynomial $f \in K[t]$ a new function

$$P_f : K \longrightarrow K : b \mapsto f(b),$$

the *polynomial function* defined by f. This way we get a map

$$P : K[t] \longrightarrow K^K : f \mapsto P_f$$

of the set the polynomials over K into the set the functions from K to K, a polynomial is maped to its polynomial function. The properties of the substitution morphism and the definition of the ring operations in K^K (see Example 6.3) show

$$P_{f+g}(b) = (f + g)(b) = f(b) + g(b) = P_f(b) + P_g(b) = (P_f + P_g)(b)$$

and

$$P_{f \cdot g}(b) = (f \cdot g)(b) = f(b) \cdot g(b) = P_f(b) \cdot P_g(b) = (P_f \cdot P_g)(b).$$

Therefore we have $P_{f+g} = P_f + P_g$ and $P_{f \cdot g} = P_f \cdot P_g$. Since moreover P_1 is the constant function 1, i.e. the 1-element of K^K, the map P is a *ring homomorphism.*

In school usually polynomials are only considered as polynomial functions, and the above ring homomorphism allows us to consider our polynomials indeed as polynomial functions. However, in general the map P is not injective, i.e. two different polynomials might define the same polynomial function. E.g. let $K = \mathbb{Z}_2$, $f = t^2+t \in K[t]$ and $g = \overline{0} \cdot t^0 \in K[t]$. Then

$$f(\overline{1}) = \overline{1} + \overline{1} = \overline{0} \quad \text{and} \quad f(\overline{0}) = \overline{0}.$$

Since $K = \{\overline{0}, \overline{1}\}$ contains only the two elements $\overline{0}$ and $\overline{1}$ the function P_f the zero function, i.e. $P_f = P_g$, even though f is not the zero polynomial, i.e. $f \neq g$.

The problem in the above example comes from the fact that K contains only finitely many elements and that thus there are polynomials besides the zero polynomial which vanish at each element in K.

If however K *is an infinite field then* P *is injective.*

To see this we only have to show that the kernel of P contains only the zero polynomial. Suppose that $0 \neq f \in \mathrm{Ker}(P)$ then P_f is the zero function, i.e. each element of K is a zero of f. By Theorem 7.42 we then get that K has at most $\deg(f)$ elements in contradiction to the assumption $|K| = \infty$.

If we work with an infinite field such as $K = \mathbb{R}$ or $K = \mathbb{C}$ then there is no need to distinguish between a polynomial and its image under the map P in K^K. □

Knowledge about the zeros of a polynomial can be helpful to see if a polynomial is irreducible or not.

Exercise 7.45 a. If K is a field and $f \in K[t]$ is a polynomial with $\deg(f) \in \{2, 3\}$. Show f is irreducible if and only if f has no zero.

b. Is $f = t^3 + 3t + 1 \in \mathbb{Z}[t]$ irreducible? If not write f as a product of irreducible polynomials.

c. Is $f_5 = t^3 + \overline{3} \cdot t + \overline{1} \in \mathbb{Z}_5[t]$ irreducible? If not write f_5 as a product of irreducible polynomials.

Exercise 7.46
Prove Theorem 7.42.

Exercise 7.47
Show that $f = t^2 + t + 1 \in \mathbb{Z}_2[t]$ is irreducible and $K = \mathbb{Z}_2[t]/\langle f \rangle$ is a field with 4 elements. Give the addition table and the multiplication table for K. What is the characteristic (see Exercise 7.21) of K? Is K isomorphic to the ring $\mathbb{Z}_4$? Is K isomorphic to the ring $\mathbb{Z}_2 \times \mathbb{Z}_2$ with componentwise operations? Consider the polynomial ring $K[x]$ over K in the indeterminate x. Is the polynomial $g = x^2 + x + \overline{1} \in K[x]$ irreducible? Has g a zero in K?

Note, in this Exercise we will write the elements $\overline{0}$ and $\overline{1}$ in $\mathbb{Z}_2$ as 0 and 1. This is sensible since the elements of $\mathbb{Z}_2[t]/\langle f \rangle$ are again residue classes and a notation like $\overline{t + \overline{1}}$ is rather clumsy.

E) Principle Ideal Domains

In the previous subsections we have that greatest common divisor exists in factorial rings and in Euclidean rings. It is an obvious question whether these concepts are somehow related, and indeed they are. The relation works via the concept of principle ideal domains.

Definition 7.48
An integral domain R is called a *principle ideal domain*, if each ideal is a principle ideal, i.e. it is generated by a single element.

In a principle ideal domain R for each ideal I there is an element $a \in I$ such that

$$I = \langle a \rangle_R = \{r \cdot a \mid r \in R\}.$$

An ideal cannot possibly be any simpler The elements in I are all multiples of the single element a.

Theorem 7.49
Each Euclidean ring is a principle ideal domain.

Proof: Let $I \trianglelefteq R$ an ideal. We have to show that $I = \langle b \rangle_R$ for a suitable element $b \in I$. Since the zero ideal is generated by 0 we may assume that $I \neq \{0\}$. If $\nu : R \setminus \{0\} \to \mathbb{N}$ is a Euclidean function of R then we can choose an element $0 \neq b \in I$ such that $\nu(b)$ is minimal. Let $0 \neq a \in I$ be an arbitrary element then there exist $q, r \in R$ such that $a = q \cdot b + r$ with $r = 0$ or $\nu(r) < \nu(b)$. Thus we have

$$r = a - q \cdot b \in I,$$

since $a \in I$ and $b \in I$. Due to the minimality of b we get necessarily $r = 0$. Therefore $a = q \cdot b \in \langle b \rangle_R$, and thus $I = \langle b \rangle_R$. □

Since we know that $\mathbb{Z}$ and the polynomial rings over fields are Euclidean we get the following Corollarys.

Corollary 7.50
$\mathbb{Z}$ is a principle ideal domain.

Corollary 7.51
If K a field then the polynomial ring $K[t]$ is a principle ideal domain.

Remark 7.52
We had already previously proved this statement for the ring of integers. By Corollary 6.30 the ideals in $\mathbb{Z}$ are additive subgroups of the additive group $\mathbb{Z}$ and by Proposition 1.39 we know that these are all generated by a single element. If you reconsider the proof of Proposition 1.39 you will see that it is identical with the proof of Theorem 7.49.

Remark 7.53
The ring

$$\mathbb{Z}\left[\frac{1+\sqrt{-19}}{2}\right] = \left\{ a + b \cdot \frac{1+\sqrt{-19}}{2} \;\middle|\; a, b \in \mathbb{Z} \right\}$$

is a principle ideal domain which is *not* Euclidean. The proof of this statement can be done with elementary methods, but it is rather technical and therefore beyond the scope of this course. □

Even though not each principle ideal domain is Euclidean we will see that also in each principle ideal domain R greatest common divisors exist. If two elements a and b in R are given then the ideal which is generated by a and b is

$$\langle a, b \rangle_R = \{ r \cdot a + s \cdot b \mid r, s \in R \}$$

and by assumption it must be generated by a single element. A generator of this ideal turns out to be a greatest common divisor of a and b.

Theorem 7.54 (Bézout Identity)
Let R be a principle ideal domain and $g, a, b \in R$. The following statements are equivalent:

a. $g \in \mathrm{GCD}(a, b)$.
b. $\langle g \rangle_R = \langle a, b \rangle_R$.

In particular, if $g \in \mathrm{GCD}(a, b)$ then there are elements $r, s \in R$ with

$$g = r \cdot a + s \cdot b. \tag{34}$$

We call (34) *a* Bézout Identity *for the greatest common divisor g of a and b.*

Proof: Let $g \in \mathrm{GCD}(a, b)$ and h a generator of the ideal $\langle a, b \rangle_R$. By Lemma 7.7

$$\langle h \rangle_R = \langle a, b \rangle_R \subseteq \langle g \rangle_R.$$

The same Lemma implies that h is a divisor of a and b. Since g is a greatest common divisor we have $h \mid g$ and therefore

$$\langle a, b \rangle_R = \langle h \rangle_R \supseteq \langle g \rangle_R.$$

Conversely the equality

$$\langle a, b \rangle_R = \langle g \rangle_R$$

implies due to Lemma 7.7 that g is a divisor of a and of b. If h is any divisor of a and of b then

$$\langle h \rangle_R \supseteq \langle a, b \rangle_R = \langle g \rangle_R,$$

so that yet again h is a divisor of g. Therefore, $g \in \mathrm{GCD}(a, b)$. □

Remark 7.55
If R is not only a principle ideal domain but a Euclidean ring then the Euclidean Algorithm can be used to compute $r, s \in R$ with $g = r \cdot a + s \cdot b$ for $g \in \mathrm{GCD}(a, b)$. For this we have to remember the q_i and the r_i in the recursion steps of the algorithm and then we have to resubstitute the results. We show this by an example.

Let $a = 8 \in \mathbb{Z}$ and $b = 3 \in \mathbb{Z}$. The Euclidean Algorithm gives:

$$\begin{aligned} 8 &= 2 \cdot 3 + 2 \\ 3 &= 1 \cdot 2 + 1 \\ 2 &= 2 \cdot 1 + 0 \end{aligned}$$

so that $1 \in \mathrm{GCD}(3, 8)$. By resubstitution then gives:

$$1 = 3 - 1 \cdot 2 = 3 - 1 \cdot (8 - 2 \cdot 3) = 3 \cdot 3 + (-1) \cdot 8.$$

With the aid of the Bézout Identity we can describe the group of units in $\mathbb{Z}_n$ find, even if n is not a prime number.

Proposition 7.56
For $0 \neq n \in \mathbb{Z}$

$$\mathbb{Z}_n^* = \{\overline{a} \mid 1 \in \mathrm{GCD}(a, n)\}$$

is the group of units of $\mathbb{Z}_n$*, i.e. a coset* $\overline{a} \in \mathbb{Z}_n$ *is invertible if and only if* 1 *is a greatest common divisor of* a *and* n*.*

Proof: Let $\overline{a} \in \mathbb{Z}_n^*$. Then there is a $\overline{b} \in \mathbb{Z}_n$ with $\overline{1} = \overline{a} \cdot \overline{b} = \overline{a \cdot b}$. Hence we have

$$a \cdot b - 1 \in n\mathbb{Z}$$

is a multiple of n. Therefore there is a $r \in R$ with

$$a \cdot b - 1 = r \cdot n,$$

and thus

$$1 = a \cdot b - r \cdot n \in \langle a, n \rangle_R \subseteq R = \langle 1 \rangle_R.$$

But then necessarily

$$\langle 1 \rangle_R = \langle a, n \rangle_R,$$

and due to Theorem 7.54 is $1 \in \mathrm{GCD}(a, n)$.

Let conversely $1 \in \mathrm{GCD}(a, n)$ then due to Theorem 7.54 there are $b, r \in R$ with

$$1 = b \cdot a + r \cdot n,$$

and therefore we have

$$\overline{1} = \overline{b \cdot a + r \cdot n} = \overline{b} \cdot \overline{a} + \overline{r} \cdot \overline{n} = \overline{b} \cdot \overline{a} \in \mathbb{Z}_n,$$

since $\overline{n} = \overline{0}$. Hence, $\overline{a} \in \mathbb{Z}_n^*$ is a unit. $\square$

Remark 7.57

The proof of Proposition 7.56 is constructive, i.e. we get an algorithm to compute the inverse of $\overline{a}$ in $\mathbb{Z}_n$, namely with aid of the Euclidean Algorithm. For $n = 8$ and $a = 3$ we determined in Remark 7.55 by the Euclidean Algorithm the following representation of 1:

$$1 = 3 \cdot 3 + (-1) \cdot 8.$$

Hence, $\overline{3}^{-1} = \overline{3} \in \mathbb{Z}_8$. □

We next want to show that each principle ideal domain is factorial so that in particular each Euclidean ring is factorial. For this we need some preparation.

Lemma 7.58

Let R be a principle ideal domain, $a \in R$ be irreducible and $b \in R \setminus \langle a \rangle_R$. Then $1 \in \mathrm{GCD}(a, b)$.

In particular there are $r, s \in R$ such that $1 = r \cdot a + s \cdot b$.

Proof: Let $g \in \mathrm{GCD}(a, b)$. It suffices to show that g is a unit. We have $a \in \langle a, b \rangle_R = \langle g \rangle_R$. Hence we have $a = c \cdot g$ for a suitable $c \in R$. Since a is irreducible either c is a unit or g is so. If c was a unit we had $\langle a \rangle_R = \langle c \cdot g \rangle_R = \langle g \rangle_R = \langle a, b \rangle_R$ in contradiction to the choice of $b \notin \langle a \rangle_R$. Therefore g must be a unit. □

Lemma 7.59

If R is a principle ideal domain then each irreducible element is prime.

Proof: Let $a \in R$ be irreducible and $a \mid b \cdot c$. Suppose $a \not| \, b$ and $a \not| \, c$, i.e. $b \in R \setminus \langle a \rangle_R$ and $c \in R \setminus \langle a \rangle_R$. Then by Lemma 7.58 there are elements $r, s, r', s' \in R$ such that

$$1 = r \cdot a + s \cdot b \quad \text{and} \quad 1 = r' \cdot a + s' \cdot c.$$

Hence we have

$$a \mid a \cdot (a \cdot r \cdot r' + r \cdot s' \cdot c + r' \cdot s \cdot b) + s \cdot s' \cdot b \cdot c = 1,$$

and a is a unit in contradiction to the assumption that a is irreducible. □

Theorem 7.60

Each principle ideal domain is factorial.

Proof: Since by Lemma 7.59 each irreducible element is prime it suffices to show that each $0 \neq a \in R \setminus R^*$ is a product of finitely many irreducible elements.

Suppose there is an element $0 \neq a_0 \in R \setminus R^*$, which is not a product of finitely many irreducible elements. Then a_0 is in particular not irreducible. Hence there are elements $0 \neq a_1, b_1 \in R \setminus R^*$ such that $a_0 = a_1 \cdot b_1$. Since a_0 is not a product of finitely many irreducible elements at least one of the two factors a_1 and b_1 isn't either. We may assume without loss of generality that for a_1 is not a product of finitely many irreducible elements. We have

$$\langle a_0 \rangle_R \subsetneqq \langle a_1 \rangle_R,$$

since b_1 is not a unit. Replace now a_0 by a_1 and reason in the same way. Going on like this we construct inductively a sequence

$$\langle a_0\rangle_R \subsetneqq \langle a_1\rangle_R \subsetneqq \langle a_2\rangle_R \subsetneqq \langle a_3\rangle_R \subsetneqq \ldots \tag{35}$$

of ideals.

We now consider their union

$$I = \bigcup_{i=0}^{\infty} \langle a_i\rangle_R$$

and we claim that this I is again an ideal. To see this note that if $b, c \in I$ then there are $i, j \in \mathbb{N}$ such that $b \in \langle a_i\rangle_R$ and $c \in \langle a_j\rangle_R$. We may assume that $i \leq j$ and therefore $\langle a_i\rangle_R \subseteq \langle a_j\rangle_R$. But then b and c both belong to $\langle a_j\rangle_R$ and since this is an ideal we also have

$$b + c \in \langle a_j\rangle_R \subseteq I.$$

Hence is I closed with respect to the addition. Moreover we have

$$r \cdot b \in \langle a_i\rangle_R \subseteq I$$

for $r \in R$. This shows that I in indeed an ideal.

Since R is a principle ideal domain I must be a principle ideal. Therefore there exists an $s \in R$ such that $I = \langle s\rangle_R$. But then we find an $i \in \mathbb{N}$ such that $s \in \langle a_i\rangle_R$ and hence

$$\langle a_{i+1}\rangle_R \subseteq I = \langle s\rangle_R \subseteq \langle a_i\rangle_R,$$

in contradiction to (35). □

Remark 7.61
The contradiction in the proof of the above theorem stems from the fact that in a principle ideal domain there is no strictly ascending chain of ideals

$$\langle a_0\rangle_R \subsetneqq \langle a_1\rangle_R \subsetneqq \langle a_2\rangle_R \subsetneqq \langle a_3\rangle_R \subsetneqq \ldots.$$

Rings with this property are called *Noetherian.* Principle ideal domain are examples of Noetherian rings. In Commutative Algebra Noetherian rings are studied further. □

Theorem 7.60 and Corollary 7.51 imply the following results.

Corollary 7.62
$\mathbb{Z}$ is factorial.

A more elaborate version of this statement is the *Fundamental Theorem of Elementary Number Theory* which follows if we take Remark 7.19 and Remark 7.20 into account since $\mathbb{Z}^* = \{1, -1\}$.

Corollary 7.63 (Fundamental Theorem of Elementary Number Theory)
For each $0 \neq z \in \mathbb{Z}$ there are uniquely determined pairwise different prime numbers $p_1, \ldots, p_k$ and uniquely determined positive integers $n_1, \ldots, n_k \in \mathbb{Z}_{>0}$ such that

$$z = \operatorname{sgn}(z) \cdot p_1^{n_1} \cdots p_k^{n_k},$$

where

$$\operatorname{sgn}(z) := \begin{cases} 1, & z > 0, \\ -1, & z < 0. \end{cases}$$

We denote by $\mathbb{P}$ *the set the prime numbers and for a prime number* $p \in \mathbb{P}$ *we introduce the notation*

$$n_p(z) = \max\left\{n \in \mathbb{N} \;\middle|\; p^n \mid z\right\}$$

which then gives

$$n_p(z) = \begin{cases} n_i, & p = p_i, \\ 0, & \textit{else} \end{cases}$$

and

$$z = \operatorname{sgn}(z) \cdot \prod_{p \in \mathbb{P}} p^{n_p(z)}.$$

Note that in the formulation of the Fundamental Theorem the product $\prod_{p\in\mathbb{P}} p^{n_p(z)}$ has at a first glance infinitely many factors. However, only finitely many of them are not equal to one, and if we use the convention that we simply forget all those factors which are one then the product is finite and well defined.

Remark 7.20 gives us the following Corollary for the computation of the greatest common divisor and lowest common multiple via prime factorisation.

Corollary 7.64
Let $a, b \in \mathbb{Z} \setminus \{0\}$ *then*

$$\gcd(a, b) = \prod_{p \in \mathbb{P}} p^{\min\{n_p(a), n_p(b)\}}$$

and

$$\operatorname{lcm}(a, b) = \prod_{p \in \mathbb{P}} p^{\max\{n_p(a), n_p(b)\}}.$$

In particular,

$$|a \cdot b| = \operatorname{lcm}(a, b) \cdot \gcd(a, b).$$

Corollary 7.65
If K *is a field then* $K[t]$ *is factorial, i.e. each polynomial in* $K[t]$ *has an essentially unique prime factorisation.*

Example 7.66
The polynomial $f = t^4 + \overline{3} \cdot t^3 + \overline{2} \in \mathbb{Z}_5[t]$ has the prime factorisation

$$f = \left(t + \overline{1}\right)^2 \cdot \left(t^2 + t + \overline{2}\right).$$

Note that $t^2 + t + \overline{2}$ is irreducible by Exercise 7.45 since the polynomial has no zero in $\mathbb{Z}_5$.

We have seen in Corollary 7.51 that the polynomial ring over a field is a principle ideal domain. The statement of the following exercise shows that the condition on K is not only sufficient but also necessary.

Exercise 7.67
For an integral domain R the following statements are equivalent:

a. R is a field.

b. $R[t]$ is a Euclidean ring.

c. $R[t]$ is a principle ideal domain.

Exercise 7.68
Let K be a field and $I \lhd K[[t]]$ be an ideal with $I \neq \{0\}$ and $I \neq K[[t]]$. Show that there is an $n \geq 1$ with $I = \langle t^n \rangle_{K[[t]]}$. Is $K[[t]]$ factorial?

F) The Chinese Remainder Theorem

In this subsection we want to answer the following question. Are there polynomials $f, g \in \mathbb{Z}[t] \setminus \mathbb{Z}^*$ such that

$$h := t^4 + 6t^3 + 17t^2 + 24t + 27 = f \cdot g,$$

i.e. is h reducible in $\mathbb{Z}[t]$? We note first that for the leading coefficient of f and g we necessarily have

$$\mathrm{lc}(f) \cdot \mathrm{lc}(g) = \mathrm{lc}(f \cdot g) = 1.$$

Thus we may assume without loss of generality that $\mathrm{lc}(f) = 1 = \mathrm{lc}(g)$.

We tackle the problem now by *reduction of the polynomial* h *modulo some prime number* p, i.e. we consider the image of h under the map

$$\phi_p : \mathbb{Z}[t] \longrightarrow \mathbb{Z}_p[t] : \sum_{k=0}^{n} a_k \cdot t^k \mapsto \sum_{k=0}^{n} \overline{a_k} \cdot t^k.$$

Since this map is a ring homomorphism the equation $h = f \cdot g$ implies

$$\phi_p(h) = \phi_p(f) \cdot \phi_p(g).$$

In the above example we consider h modulo the prime numbers 2 and 7, and we get

$$\phi_2(h) = t^4 + \overline{6} \cdot t^3 + \overline{17} \cdot t^2 + \overline{24} \cdot t + \overline{27} = t^4 + t^2 + \overline{1} = \left(t^2 + t + \overline{1}\right)^2 \in \mathbb{Z}_2[t] \tag{36}$$

and

$$\begin{aligned}\phi_7(h) = t^4 + \overline{6} \cdot t^3 + \overline{17} \cdot t^2 + \overline{24} \cdot t + \overline{27}& \\ = t^4 + \overline{6} \cdot t^3 + \overline{3} \cdot t^2 + \overline{3} \cdot t + \overline{6}& \\ = \left(t^2 + \overline{5} \cdot t + \overline{2}\right) \cdot \left(t^2 + t + \overline{3}\right) \in \mathbb{Z}_7[t].&\end{aligned}$$

The factorisation of $\phi_2(h)$ in $\mathbb{Z}_2[t]$ and of $\phi_7(h)$ in $\mathbb{Z}_7[t]$ can be computed by considering all products of all pairs of two polynomials of degree at most three whose degrees add up to four. Since there are only finitely many this can be done in finite time. The factors of $\phi_2(h)$ and of $\phi_7(h)$ which we found that way are irreducible by Exercise 7.45 since their degree is two and they have no zero in $\mathbb{Z}_2$ respectively $\mathbb{Z}_7$. The latter can easily be checked.

Thus if there are polynomials f and g as above then their degree must necessarily be two, i.e.

$$f = t^2 + b_1 \cdot t + b_0 \quad \text{and} \quad g = t^2 + c_1 \cdot t + c_0,$$

and moreover their reduction modulo 2 respectively modulo 7 satisfies

$$\phi_2(f) = \phi_2(g) = t^2 + t + \overline{1}$$

respectively

$$\phi_7(f) = t^2 + \overline{5} \cdot t + \overline{2} \quad \text{and} \quad \phi_7(g) = t^2 + t + \overline{3}.$$

We are thus looking for integers $b_0, b_1, c_0, c_1 \in \mathbb{Z}$ which satisfy the following system of congruence equations:

$$\begin{aligned} b_0 &\equiv 1 \pmod 2 \\ b_0 &\equiv 2 \pmod 7 \end{aligned} \tag{37}$$

$$\begin{aligned} b_1 &\equiv 1 \pmod 2 \\ b_1 &\equiv 5 \pmod 7 \end{aligned} \tag{38}$$

$$\begin{aligned} c_0 &\equiv 1 \pmod 2 \\ c_0 &\equiv 3 \pmod 7 \end{aligned} \tag{39}$$

$$\begin{aligned} c_1 &\equiv 1 \pmod 2 \\ c_1 &\equiv 3 \pmod 7 \end{aligned} \tag{40}$$

Can we solve a system of simultaneous congruence equations like (37)? The answer to this question gives the Chinese Remainder Theorem, an algorithm for solving such systems that was already known in China in the third century.

The following Lemmata are important steps in the proof of the Chinese Remainder Theorem.

Lemma 7.69
Let $n_1, \ldots, n_r \in \mathbb{Z} \setminus \{0\}$ be pairwise coprime and let $N_i = \frac{n_1 \cdots n_r}{n_i}$.

Then n_i and N_i are coprime and $\overline{N_i} \in \mathbb{Z}_{n_i}^*$ for $i \in \{1, \ldots, n\}$.

Proof: Let $i \in \{1, \ldots, r\}$ be given. For $j \neq i$ the integers n_i and n_j are coprime. This means that $1 \in \mathrm{GCD}(n_i, n_j)$, and due to the Bézout Identity there exist integers $s_j, r_j \in R$ such that

$$1 = n_i \cdot r_j + n_j \cdot s_j.$$

If j passes through all indices from 1 to r except for i then we can write the integer 1 in the following way as a product of $r - 1$ factors:

$$1 = \prod_{j \neq i} 1 = \prod_{j \neq i} (n_i \cdot r_j + n_j \cdot s_j). \tag{41}$$

Expanding the product on the right hand side we get a sum in which all but one summand obviously contain the integer n_i as a factor. The only summand which is not a multiple of n_i is

$$\prod_{j\neq i}(n_j \cdot s_j) = N_i \cdot \prod_{j\neq i} s_j.$$

If we split of n_i of those summands which are divisible by n_i we can rewrite the above Equation (41) as:

$$1 = n_i \cdot z + N_i \cdot \prod_{j\neq i} s_j \in \langle n_i, N_i\rangle_{\mathbb{Z}}$$

where $z \in \mathbb{Z}$ is a suitable integer. But then we have $\langle n_i, N_i\rangle_{\mathbb{Z}} = \langle 1\rangle_{\mathbb{Z}}$ and due to Theorem 7.54 $1 \in \mathrm{GCD}(n_i, N_i)$, i.e. n_i and N_i are coprime. Proposition 7.56 finally implies that N_i is a unit in $\mathbb{Z}_{n_i}$. □

Lemma 7.70
If $n_1, \ldots, n_r \in \mathbb{Z}\backslash\{0\}$ *are pairwise coprime and* $a \in \mathbb{Z}\backslash\{0\}$ *with* $n_i \mid a$ *for* $i = 1, \ldots, r$ *we have:*

$$n_1 \cdots n_r \mid a.$$

Proof: We do the proof by induction on r where the claim for $r = 1$ is obviously fulfilled. We can therefore assume that $r \geq 2$.

With the notation of Lemma 7.69 we have then by induction assumption

$$N_r = n_1 \cdots n_{r-1} \mid a.$$

Hence there are integers $b, c \in \mathbb{Z}$ with $a = n_r \cdot b$ and $a = N_r \cdot c$. Since by Lemma 7.69 moreover n_r and N_r are coprime the Bézout Identity gives integers $x, y \in \mathbb{Z}$ with

$$x \cdot n_r + y \cdot N_r = 1.$$

If we combine the three equations we get:

$$\begin{aligned} a = a \cdot (x \cdot n_r + y \cdot N_r) &= a \cdot x \cdot n_r + a \cdot y \cdot N_r \\ &= N_r \cdot c \cdot x \cdot n_r + n_r \cdot b \cdot y \cdot N_r = n_r \cdot N_r \cdot (c \cdot x + b \cdot y). \end{aligned}$$

Hence a is a multiple of $N_r \cdot n_r = n_1 \cdots n_r$. □

Theorem 7.71 (Chinese Remainder Theorem)
Let $n_1, \ldots, n_r \in \mathbb{Z} \setminus \{0\}$ *be pairwise coprime,* $N = n_1 \cdots n_r$ *and* $N_i = \frac{N}{n_i}$.

a. *For arbitrary integers* $a_1, \ldots, a_r \in \mathbb{Z}$ *there exists a solution* $x \in \mathbb{Z}$ *of the simultaneous* system of congruence equations

$$\begin{aligned} x &\equiv a_1 \pmod{n_1}, \\ x &\equiv a_2 \pmod{n_2}, \\ &\vdots \\ x &\equiv a_r \pmod{n_r}. \end{aligned} \tag{42}$$

b. *If* $\overline{x_i} = \overline{N_i}^{-1} \in \mathbb{Z}_{n_i}$ *for* $i = 1, \ldots, r$ *then*

$$x' = \sum_{i=1}^{r} N_i \cdot x_i \cdot a_i \in \mathbb{Z} \tag{43}$$

is a solution of (42).

c. $x'' \in \mathbb{Z}$ *is a solution of* (42) *if and only if if* x'' *differs from* x' *only by a multiple of* N*. In particular, the solution of* (42) *is uniquely determined modulo* N*.*

Proof: We show first that x' is a solution of (42) and we prove thereby a. and b.. By Lemma 7.69 for $i = 1, \ldots, r$ there exists an $x_i \in \mathbb{Z}$ with $\overline{x_i} = \overline{N_i}^{-1} \in \mathbb{Z}_{n_i}$. We thus can consider

$$x' := \sum_{j=1}^{r} N_j \cdot x_j \cdot a_j.$$

Since $n_i \mid N_j$ for $j \neq i$ in $\mathbb{Z}_{n_i}$ we have the equation

$$\overline{x'} = \sum_{j=1}^{r} \overline{N_j} \cdot \overline{x_j} \cdot \overline{a_j} = \overline{N_i} \cdot \overline{x_i} \cdot \overline{a_i} = \overline{a_i} \in \mathbb{Z}_{n_i},$$

i.e.

$$x' \equiv a_i \ (\mathrm{mod}\ n_i).$$

It remains to show therefore that $x' + N\mathbb{Z}$ is the set the solutions of (42). Let $x'' \in \mathbb{Z}$ be an arbitrary solution of (42). Then we have for $i = 1, \ldots, r$

$$x' - a_i, x'' - a_i \in n_i\mathbb{Z}.$$

Therefore we get $x' - x'' \in n_i\mathbb{Z}$, i.e. $n_i \mid (x' - x'')$, for all $i = 1, \ldots, r$. Lemma 7.70 implies then $N \mid (x' - x'')$, d. h. $x' - x'' \in N\mathbb{Z}$ and therefore

$$x' \equiv x'' \ (\mathrm{mod}\ N).$$

If conversely $x'' = x' + N \cdot z$ for a $z \in \mathbb{Z}$ then $N \mid x' - x''$ and therefore $n_i \mid x' - x''$ for all $i = 1, \ldots, r$. Since we know already that $x' \equiv a_i \ (\mathrm{mod}\ n_i)$, i.e. $n_i \mid x' - a_i$, we deduce

$$n_i \mid \big((x' - a_i) - (x' - x'')\big) = (x'' - a_i),$$

i.e. $x'' \equiv a_i \ (\mathrm{mod}\ n_i)$ for all $i = 1, \ldots, r$. Therefore x'' is also a solution of (42). □

Remark 7.72

Since we can compute the inverse of $\overline{N_i}$ in $\mathbb{Z}_{n_i}$ with aid of the Euclidean Algorithm (see Remark 7.57) we are indeed able to solve a system of congruence equations of the form (42) with aid the formula (43).

In applications the n_i are usually pairwise different prime numbers as in our introductory example.

We can formulate the statement of the Chinese Remainder Theorem in a more algebraic way, if we consider the Cartesian product

$$\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \ldots \times \mathbb{Z}_{n_r}$$

with componentwise addition and multiplication as a commutative ring with one. The algebraic formulation will be most important for the lecture Elementary Number Theory.

In the following Corollary we denote the coset of x in $\mathbb{Z}_m$ by $\overline{x}_m$ instead of $\overline{x}$ in order to make clear what the modulus is.

Corollary 7.73 (Chinese Remainder Theorem)
Let $n_1, \ldots, n_r \in \mathbb{Z}_{>0}$ be pairwise coprime positive numbers then the map

$$\overline{\alpha} : \mathbb{Z}_{n_1 \cdots n_r} \longrightarrow \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r} : \overline{x}_{n_1 \cdots n_r} \mapsto \left(\overline{x}_{n_1}, \ldots, \overline{x}_{n_r}\right)$$

is an isomorphism of commutative rings with one.

Moreover, the induced map

$$\mathbb{Z}^*_{n_1 \cdots n_r} \longrightarrow \mathbb{Z}^*_{n_1} \times \ldots \times \mathbb{Z}^*_{n_r} : \overline{x}_{n_1 \cdots n_r} \mapsto \left(\overline{x}_{n_1}, \ldots, \overline{x}_{n_r}\right)$$

is an isomorphism of the groups of units of the rings.

Proof: The map

$$\alpha : \mathbb{Z} \longrightarrow \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r} : x \mapsto \left(\overline{x}_{n_1}, \ldots, \overline{x}_{n_r}\right)$$

is obviously a ring homomorphism. The Chinese Remainder Theorem 7.71 then implies that α surjective with

$$\mathrm{Ker}(\alpha) = \langle n_1 \cdots n_r \rangle_{\mathbb{Z}}.$$

The first claim therefore follows by the Homomorphism Theorem 4.50.

Since an isomorphism of rings maps units to units we have

$$\overline{\alpha}\big(\mathbb{Z}^*_{n_1 \cdots n_r}\big) = \left(\mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_r}\right)^* = \mathbb{Z}^*_{n_1} \times \ldots \times \mathbb{Z}^*_{n_r},$$

where the last equality follows from Example 6.4. Since the map $\overline{\alpha}$ respects the multiplication it induces an isomorphism of the multiplicative groups. $\square$

Example 7.74
We want to solve the systems of congruence equations (37), (38), (39) and (40). The first one has the form:

$$\begin{aligned} b_0 &\equiv 1 \pmod 2 \\ b_0 &\equiv 2 \pmod 7 \end{aligned}$$

In the notation of the Chinese Remainder Theorem $n_1 = N_2 = 2$, $n_2 = N_1 = 7$, $a_1 = 1$ and $a_2 = 2$. Even without applying the Euclidean Algorithm we see that

$$1 = 4 \cdot 2 + (-1) \cdot 7.$$

Hence we have $\overline{x_1} = \overline{1} = \overline{-1} = \overline{N_1}^{-1} \in \mathbb{Z}_2$ and $\overline{x_2} = \overline{4} = \overline{N_2}^{-1} \in \mathbb{Z}_7$. The solution b_0 can thus up to a multiple of $N = 2 \cdot 7 = 14$ be written as

$$b_0 \equiv x_1 \cdot N_1 \cdot a_1 + x_2 \cdot N_2 \cdot a_2 = 1 \cdot 7 \cdot 1 + 4 \cdot 2 \cdot 2 = 23 \equiv 9 \pmod{14}.$$

For the remaining three systems of congruence equations the integers $\overline{n_i}$, $\overline{N_i}$ and $\overline{x_i}$ stay unchanged and only the a_i vary. We thus can compute the solution modulo $N = 14$ without further ado:

$$b_1 \equiv x_1 \cdot N_1 \cdot 1 + x_2 \cdot N_2 \cdot 5 = 1 \cdot 7 \cdot 1 + 4 \cdot 2 \cdot 5 = 47 \equiv 5 \pmod{14},$$

$$c_0 \equiv x_1 \cdot N_1 \cdot 1 + x_2 \cdot N_2 \cdot 3 = 1 \cdot 7 \cdot 1 + 4 \cdot 2 \cdot 3 = 31 \equiv 3 \pmod{14}$$

and

$$c_1 \equiv x_1 \cdot N_1 \cdot 1 + x_2 \cdot N_2 \cdot 5 = 1 \cdot 7 \cdot 1 + 4 \cdot 2 \cdot 3 = 31 \equiv 3 \pmod{14}.$$

If we knew by some extra considerations that the coefficients of f and g had to lie between 0 and 13 we could determine the polynomials f and g absolutely as

$$f = t^2 + b_1 \cdot t + b_0 = t^2 + 5t + 9$$

and

$$g = t^2 + c_1 \cdot t + c_0 = t^2 + 3t + 3.$$

We do not have these additional information, however, we can just test our result, and we get indeed

$$f \cdot g = (t^2 + 5t + 9) \cdot (t^2 + 3t + 3) = t^4 + 6t^3 + 17t^2 + 24t + 27 = h.$$

□

Remark 7.75

There are indeed results on the growth of the coefficients in a factorisation of a polynomial in $\mathbb{Z}[t]$ in terms of the coefficients of the polynomial itself. If one then chooses enough prime numbers such that the upper bound for the coefficients is smaller than their product one can decompose the polynomial in $\mathbb{Z}[t]$ into irreducible factors in the described way. Furthermore, there is a result which states that a polynomial which is irreducible in $\mathbb{Z}[t]$ is also irreducible in $\mathbb{Q}[t]$. This finally leads to an algorithm for computing a prime factorisation of polynomials in $\mathbb{Q}[t]$.

Note that there exists no such algorithm for polynomials in $\mathbb{R}[t]$ or $\mathbb{C}[t]$, which is a major reason for the necessity of numerical algorithms. □

We want to conclude this section by a somewhat more elaborate example for the Chinese Remainder Theorem.

Example 7.76

Consider the following system of congruence equations:

$$\begin{aligned} x &\equiv a_1 = 1 \pmod{2}, \\ x &\equiv a_2 = 2 \pmod{3}, \\ x &\equiv a_3 = 4 \pmod{7}. \end{aligned}$$

The integers $n_1 = 2, n_2 = 3, n_3 = 7$ are pairwise coprime, and $N = 2 \cdot 3 \cdot 7 = 42$, $N_1 = 21$, $N_2 = 14$ and $N_3 = 6$.

The computation of the inverse of $\overline{N_i}$ in $\mathbb{Z}_{n_i}$ is done with aid of the Euclidean Algorithm. Since n_i and N_i are coprime we have due to the Bézout Identity

$$x_i N_i + y_i n_i = 1$$

for suitable $x_i \in \mathbb{Z}$ (and $y_i \in \mathbb{Z}$ which is of no interest to us here):

$$\overline{x_1} = \overline{21}^{-1} = \overline{1}^{-1} = \overline{1} \in \mathbb{Z}_2,$$

$$\overline{x_2} = \overline{14}^{-1} = \overline{2}^{-1} = \overline{2} \in \mathbb{Z}_3,$$

and

$$\overline{x_3} = \overline{6}^{-1} = \overline{6} \in \mathbb{Z}_7.$$

We thus get:

$$\begin{aligned} x &\equiv N_1 \cdot x_1 \cdot a_1 + N_2 \cdot x_2 \cdot a_2 + N_3 \cdot x_3 \cdot a_3 \\ &= 21 \cdot 1 \cdot 1 + 14 \cdot 2 \cdot 2 + 6 \cdot 4 \cdot 6 = 221 \equiv 11 \pmod{42}. \end{aligned}$$

Therefore $x = 11$ is modulo 42 the uniquely determined solution, and the set of all solutions is

$$\{11 + z \cdot 42 \mid z \in \mathbb{Z}\}.$$

□

Remark 7.77
The assumption of the Chinese Remainder Theorem that the n_i are pairwise coprime is not only necessary for our proof. Without the assumption the statement itself is in general wrong as following example shows: $n_1 = 2$, $n_2 = 4$, $a_1 = 0$, $a_2 = 1$; the equation $x \equiv a_1 \pmod 2$ implies that x is an even integer while the equation $x \equiv a_2 \pmod 4$ can only be satisfied by odd integers. Therefore there cannot exist any integer x which satisfies both congruence equations simultaneously. This is due to the fact the moduli $n_1 = 2$ and $n_2 = 4$ are not coprime.

The last result of these lecture notes is devoted to the question how many prime numbers there exist. This is important since by the Fundamental Theorem of Elementary Number Theory the prime numbers are the basic building blocks of all integers.

Theorem 7.78 (Euclid)
There are infinitely man prime numbers in $\mathbb{Z}$.

Proof: Since 2 is a prime number there exists at least one prime number. Suppose now there were only finitely many prime numbers $p_1, \ldots, p_r \in \mathbb{Z}$, and consider the integer

$$z = p_1 \cdots p_r + 1 > 1.$$

By the Fundamental Theorem of Elementary Number Theory z must have a prime factorisation which means in particular that some prime number divides z, i.e. there is some i such that $p_i \mid z$. But then we get

$$p_i \mid z - p_1 \cdots p_r = 1$$

which is only possible if p_i is a unit. This however is in contradiction to the assumption that p_i is a prime number.

This shows that there are infinitely many prime numbers. $\square$

Index

Literatur

[GK00] Gert-Martin Greuel and Thomas Keilen, *Lineare Algebra I & II*, Vorlesungsskript, FB Mathematik, Universität Kaiserslautern, 2000.

[Hum96] John F. Humphreys, *A course in group theory*, OUP, Oxford, 1996.

[Kei01] Thomas Keilen, *Endliche Gruppen*, Fachbereich Mathematik, Universität Kaiserslautern, Jan. 2001, Proseminarskript, 3. Auflage, http://www.mathematik.uni-kl.de/˜wwwagag/download/scripts/Endliche.Gruppen.ps.gz.

[Kei02] Thomas Keilen, *Algebra I*, Mathematics Institute, University of Warwick, Oct. 2002, http://www.mathematik.uni-kl.de/˜keilen/download/Lehre/ALGWS02/algebra.ps.gz.

[Lan97] Serge Lang, *Algebraische Strukturen*, Springer, 1997.

[Sie81] Helmut Siemon, *Anwendungen der elementaren Gruppentheorie: in Zahlentheorie und Kombinatorik*, Klett Studienbücher, Klett, 1981.

[Ver75] J. Verhoeff, *Error detecting decimal codes*, Mathematical Centre Tracts, no. 29, Mathematisch Centrum Amsterdam, 1975.