Lecture Notes for Abstract Algebra (MS-C1081)

Milo Orlich

February 22, 2023

Contents

These notes are written for the 2022/2023 implementation of Abstract Algebra (MS-C1081). They are heavily based on the notes by Kaie Kubjas for the 2021/2022 implementation. The students are encouraged to deepen their understanding by consulting textbooks and other material, as suggested in MyCourses.

If you spot typos or just wish to give feedback, please send a message to

`milo.orlich@aalto.fi`

Notation.

- The symbol ":=" means that the left-hand side is an abbreviation of the right-hand side.
- Given two sets X and Y, the ***Cartesian product*** of X and Y is the set

$$X \times Y := \{(x,y) \mid x \in X, y \in Y\}.$$

- The symbol "$\subseteq$" stands for inclusion of sets, with possible equality.
- $\#X$ is the cardinality of X, i.e., the number of elements in X.
- In this course, the set $\mathbb{N}$ of natural numbers contains 0.
- $\mathcal{U}(X)$ is the group of invertible elements
- The notation $H \lhd G$ means that H is a normal subgroup of G.

Chapter 1

Groups

A square matrix M is *invertible* if there exists a matrix N such that $MN = I$ and $NM = I$, where I is the identity matrix of the appropriate size, and in this case we say that N is an *inverse matrix of* M. We learn early in the study of matrices that if N and N' both are inverse matrices of M, then $N = N'$. This is easily proven by the associativity of the matrix product and the fact that any matrix is equal to the product of itself by the identity matrix:

$$N = NI = N(MN') = (NM)N' = IN' = N'.$$

Therefore we can simply denote the inverse by M^{-1}, since it is uniquely determined by M.

When we first learn about integers, we learn that any natural number m has an *opposite number* "$-m$". We learn that $m + (-m) = 0$,[1] and we usually do not even question the uniqueness of the opposite. But indeed it is uniquely determined: if n and n' are two integers such that $m + n = 0$ and $m + n' = 0$, then we can write

$$n = n + 0 = n + (m + n') = (n + m) + n' = 0 + n' = n',$$

where we used the associativity of the sum and the fact that any integer is equal to the sum of itself and zero.

You should notice that these calculations look very similar, and indeed they are special instances of the general fact that the inverse for an associative operation, if it exists, is uniquely determined (Proposition 1.12). A key point in abstract algebra (more generally, in pure mathematics) is to realize that we do not need very special properties of invertible matrices or integers in order to prove the statements above. We simply needed the fact that there is a neutral element (the identity matrix and number zero, respectively) and that the operation under consideration in each case is associative. The existence of a neutral element and associativity of the operation will turn out to be some of the fundamental properties of a group. Proving things in general starting from few axioms allows to

- save energy, by proving only once a general result that can be specialized later to many individual little theories;
- achieve a better, unified understanding of the general theory, from a higher point of view. The special methods of each single theory do not lose their use or interest, as they are still fundamental in actual computations and so on.

The concept of a group occurs in most parts of pure and applied mathematics. Group theory can be seen as the "study of symmetry", which is also related to physics, architecture, chemistry, etc.

[1] And since we also learn that the sum of integers is commutative, this means that $(-m) + m = 0$, too.

1.1 Binary operations and rudiments of groups

Definition 1.1. A ***binary operation*** $*$ on a set X is a map from $X \times X$ to X. We will denote the value $*(a,b)$ by $a * b$.

Examples 1.2. For instance

$$\begin{aligned} \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} &\longrightarrow \mathbb{Z}_{>0} \\ (a,b) &\longmapsto a^b \end{aligned} \qquad \begin{aligned} \mathbb{Z} \times \mathbb{Z} &\longrightarrow \mathbb{Z} \\ (a,b) &\longmapsto a-b \end{aligned} \qquad \begin{aligned} \{0,1\} \times \{0,1\} &\longrightarrow \{0,1\} \\ (x,y) &\longmapsto \min\{x,y\} \end{aligned}$$

are binary operations, with $X = \mathbb{Z}_{>0}$, $X = \mathbb{Z}$ and $X = \{0,1\}$, respectively.

Definition 1.3. Let $*$ be a binary operation on a set X.

- We say $*$ is ***associative*** if

$$(a * b) * c = a * (b * c) \qquad \text{for all } a,b,c \in X.$$

- We say $*$ is ***commutative*** if

$$a * b = b * a \qquad \text{for all } a,b \in X.$$

Examples 1.4.

1. Of the first two operations in Examples 1.2, neither is associative or commutative:

$$(2^1)^3 \neq 2^{(1^3)}, \qquad 2^1 \neq 1^2, \qquad (1-2)-3 \neq 1-(2-3), \qquad 1-2 \neq 2-1.$$

   On the other hand, the third operation is both associative and commutative.

2. Given a set X, denote $X^X := \{\text{functions } X \to X\}$. The operation

$$\begin{aligned} X^X \times X^X &\longrightarrow X^X \\ (f,g) &\longmapsto g \circ f, \end{aligned}$$

   that is, the usual composition of functions, is always associative. However, it is in general not commutative. Exercise.

3. For an operation that is commutative but not associative, see Examples 1.7.

4. The ordinary sum $+\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and product $\cdot\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ of real numbers are both associative and commutative.

1.1.1 Cayley tables

If the set is finite, it is common to visualize a binary operation on a table.

Definition 1.5. Given a finite set X and an operation $*$ on X, the ***Cayley table of*** $*$ is the array with rows and columns indexed by the elements of X, and where one writes $a*b$ in the row corresponding to element a and the column corresponding to b.

Remark 1.6. One can easily detect commutativity from a Cayley table: it occurs exactly when the table is mirrored about the main diagonal (the diagonal “with slope -1”).

Examples 1.7. 1. Consider the following operation on the set $X = \{1,2,3\}$:

$*$	1	2	3
1	2	1	1
2	3	1	2
3	1	3	1

$$\begin{array}{lll} 1*1=2 & 1*2=1 & 1*3=1 \\ 2*1=3 & 2*2=1 & 2*3=2 \\ 3*1=1 & 3*2=3 & 3*3=1. \end{array}$$

The Cayley table is displayed on the left. This operation is not commutative.

2. Say we want to define an operation that is commutative but not associative, on the set $X = \{1,2,3\}$. Let's say that we want $1*(2*3) \neq (1*2)*3$. If we require that $1*2 = 1$, $1*3 = 2$ and $2*3 = 2$, this is enough to achieve the goal. This information is recorded in the Cayley table as follows:

$*$	1	2	3
1		1	2
2			2.
3			

But we also want to make this operation commutative, which means that three more entries have to be filled as follows:

$*$	1	2	3
1		1	2
2	1		2
3	2	2.	

And lastly, regardless of how we fill the remaining three empty entries, the resulting operation will still be commutative and not associative.

To be pedantic, a Cayley table depends on the chosen order of the elements of the set on which the operation is defined. By changing that order, one permutes rows and columns of the table accordingly.

1.1.2 Neutral elements and inverses

Definition 1.8. Let $*$ be a binary operation on a set X. We say that an element $e \in X$ is a ***neutral element for*** $*$ if

$$e*a = a*e = a \qquad \text{for all } a \in X.$$

Proposition 1.9. *Given a binary operation $*$, there is at most one neutral element for $*$. More precisely, if e_1 and e_2 are both neutral elements for $*$, then $e_1 = e_2$.*

Proof. Since e_1 is a neutral element, we have $e_2 = e_1 * e_2$. And since e_2 is a neutral element, we have $e_1 = e_1 * e_2$. Hence $e_1 = e_2$. □

Examples 1.10. 1. The first two operations in Examples 1.2 do not have a neutral element, and the third operation has number 1 as neutral element.

2. The integers may for instance be equipped with the ordinary sum, and the neutral element with respect to that is number 0, or with ordinary multiplication, and the neutral element with respect to that is number 1.

3. The set $M_{n,m}(\mathbb{R})$ of $n \times m$ real matrices may be equipped with the ordinary matrix sum, and the neutral element with respect to that is the zero matrix, i.e., the $n \times m$ matrix filled with zeros. (This operation is associative and commutative.)

4. The set $M_n(\mathbb{R})$ of $n \times n$ real matrices may be equipped with the ordinary matrix product, ant the neutral element with respect to that is the identity matrix, with ones on the diagonal and zeros elsewhere. (This operation is associative but not commutative.)

Definition 1.11. Let $*$ be a binary operation on X, with neutral element e. We say that an element $x \in X$ is ***invertible (with respect to** $*$**)*** if there exists $y \in X$ such that

$$x * y = y * x = e.$$

Such an element y is called an ***inverse of*** x. In virtue of the following result, for an *associative* operation the inverse of x is unique, and in this case we denote it by x^{-1}.

Proposition 1.12. *Let $*$ be an* associative *binary operation on X, with neutral element e. If both y and y' are inverses of the same element x with respect to $*$, then $y = y'$.*

Proof. We have

$$y = y * e = y * (x * y') = (y * x) * y' = e * y' = y'.$$

□

Note that the calculation in the above proof is an abstraction of what happened at the very beginning of this chapter.

Examples 1.13. 1. The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ may all be equipped with their ordinary sum. The inverse of x with respect to the sum is usually called the *opposite* of x, and denoted $-x$. All elements in all of the sets mentioned here have an inverse with respect to the sum.

2. The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ may all be equipped with their ordinary product. The inverse of x with respect to the product is usually denoted $\frac{1}{x}$. In $\mathbb{Z}$, the only invertible elements with respect to the product are 1 and -1, whereas in $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ all elements except for zero are invertible with respect to the product.

3. Consider a set X and equip $X^X := \{\text{functions } X \to X\}$ with function composition. The invertible elements with respect to composition are the bijective functions $X \to X$. The inverse of f is usually denoted f^{-1}.

Remark 1.14. When we use the symbols $+$ and $\cdot$ to denote binary operations, and refer to them as "sum" or "product", respectively, we then say "additive inverse" or "multiplicative inverse", respectively.

Proposition 1.15. *Let $*$ be an associative binary operation, with neutral element e. Let x and y be invertible elements. Then $x * y$ is invertible, and more precisely*

$$(x * y)^{-1} = y^{-1} * x^{-1}.$$

Proof. We have

$$(x * y) * (y^{-1} * x^{-1}) = x * (y * y^{-1}) * x^{-1} = x * e * x^{-1} = x * x^{-1} = e,$$

and similarly $(y^{-1} * x^{-1}) * (x * y) = e$. □

1.1.3 Groups – Definition, first examples and basic results

Definition 1.16. A ***group*** $(G, *)$ consists of a set G equipped with a binary operation $*$ on G that satisfies the following conditions:

1. $*$ is associative,
2. $*$ has a neutral element, often called ***identity element*** (unique by Proposition 1.9),
3. every element has an inverse (unique by Proposition 1.12).[2]

Examples 1.17. 1. Each of $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$ is a group, with identity element 0.

2. Each of $(\mathbb{Q} \setminus \{0\}, \cdot)$, $(\mathbb{R} \setminus \{0\}, \cdot)$, $(\mathbb{C} \setminus \{0\}, \cdot)$ is a group, with identity element 1.
3. None of $(\mathbb{N}, +)$, $(\mathbb{N}, \cdot)$, $(\mathbb{Z}, \cdot)$, $(\mathbb{Z} \setminus \{0\}, \cdot)$, $(\mathbb{R}, \cdot)$ is a group, because there are some elements with no inverses. The other two conditions in the group definition are satisfied.
4. Denoting $2\mathbb{Z}$ the set of even integers, $(2\mathbb{Z}, +)$ is a group. On the other hand, $(2\mathbb{Z}, \cdot)$ is not a group, because it does not have a neutral element (but the other two conditions are satisfied).
5. Denote by $M_{n,m}(\mathbb{R})$ the set of $m \times n$ matrices with entries in $\mathbb{R}$. Then $(M_{m,n}(\mathbb{R}), +)$ is a group. We write $M_n(\mathbb{R}) := M_{n,n}(\mathbb{R})$.
6. Denote by $\mathrm{GL}_n(\mathbb{R})$ the set of matrices in $M_n(\mathbb{R})$ which are invertible with respect to the usual matrix product. Then $(\mathrm{GL}_n(\mathbb{R}), \cdot)$ is a group, and $(M_n(\mathbb{R}), \cdot)$ is not.
7. On the set $\mathbb{R}^{\mathbb{R}} := \{\text{functions } \mathbb{R} \to \mathbb{R}\}$ we may define a "sum" (called *pointwise sum*): for $f, g \in \mathbb{R}^{\mathbb{R}}$, the function $f + g$ is defined by setting
$$(f+g)(x) := f(x) + g(x) \qquad \text{for all } x \in \mathbb{R},$$
where the sum on the right-hand side is the one for real numbers. Then $(\mathbb{R}^{\mathbb{R}}, +)$ is a group: the associativity of $+$ follows from the associativity of the sum of real numbers, the identity element is the function constantly equal to zero, and for any $f \in \mathbb{R}^{\mathbb{R}}$, the inverse is the function $-f$ defined by $(-f)(x) := -(f(x))$.
8. One may generalize the construction in the previous item by defining an analogous operation on $G^X := \{\text{functions } X \to G\}$, where X is a set and G is a group. Again G^X inherits its group structure from G.

Remark 1.18. Sometimes we will be sloppy regarding the notation and refer to a group G instead of $(G, *)$ with the understanding that there is an underlying binary operation on the set G. If we need to specify the binary operation $*$, the we say "the group G under $*$".

Definition 1.19. A group $(G, *)$ is called an ***abelian group*** if the operation $*$ is commutative, that is, if
$$a * b = b * a \quad \text{for all } a, b \in G.$$

Example 1.20. Among the groups listed in Examples 1.17, those in items 1, 2, 5 and 7 are abelian. What about the other ones?

[2] If we assume only the first condition in Definition 1.16 to hold, we call $(G, *)$ a *semigroup*. If the first two conditions hold, we call $(G, *)$ a *monoid*. These algebraic structures will not be discussed in this course.

Remark 1.21. Very often, the operation symbol in an *abelian* group is chosen to be $+$ and in this case the identity is called 0. Otherwise the identity element e is often denoted 1.

Remark 1.22. Often one does not use a special symbol $*$ to denote a binary operation different from addition or multiplication. Instead, one uses either $a+b$ or $a \cdot b$. Moreover, the multiplication is usually written by juxtaposition without a dot, i.e., we write ab. The sum notation is usually used only for commutative operations. From now on, we will mostly use this convention.

Proposition 1.23. *Let $(G,\cdot)$ be a group. Then*

$$a \cdot b = a \cdot c \text{ implies } b = c \qquad \text{and} \qquad b \cdot a = c \cdot a \text{ implies } b = c,$$

for all $a,b,c \in G$. These are called ***left and right cancellation laws****, respectively.*

Proof. Exercise. □

Corollary 1.24. *Let $(G,\cdot)$ be a group and $a,b \in G$. Then the linear equation system*

$$\begin{cases} a \cdot x = b \\ y \cdot a = b \end{cases}$$

has a unique solution (x,y) in $G \times G$.

Proof. Exercise. □

1.1.4 Important examples: $\mathbb{Z}_n$, symmetric group S_n, dihedral group D_n

Here we introduce some of the most important groups considered in this course. First of all, a useful lemma:

Lemma 1.25. *Let $*$ be an associative binary operation on X, with neutral element e. Define the set*

$$\mathcal{U}(X) := \{x \in X \mid x \text{ is invertible with respect to } *\}.$$

Then $(\mathcal{U}(X),*)$, where we simply denote by $*$ the restriction of $*$ to $\mathcal{U}(X) \times \mathcal{U}(X)$, is a group.

Proof. By Proposition 1.15, if x and y are in $\mathcal{U}(X)$, then $x * y \in \mathcal{U}(X)$, so that $*$ is a binary operation on $\mathcal{U}(X)$. Moreover, the restriction of an associative operation to a subset is still associative. The neutral element is invertible, with inverse equal to itself: $e * e = e$. And e is the neutral element on $\mathcal{U}(X)$, too. Lastly, if $x \in \mathcal{U}(X)$, then x is invertible by definition of $\mathcal{U}(X)$, which means that x^{-1} exists, but in fact x^{-1} is an element of $\mathcal{U}(X)$, since x is its inverse. □

Operations in $\mathbb{Z}_n$, a.k.a. operations modulo n

Some familiarity with operations modulo n is assumed. You should have already seen parts of what follows in some course such as Foundations of Discrete Math. Nevertheless, a little summary follows.

Theorem 1.26 (Euclid's theorem). *For any two integers m and n, we may write uniquely*

$$m = qn + r, \qquad \text{with } q,r \in \mathbb{Z} \text{ and } 0 \leq r < n.$$

The number r is called the ***remainder of** m **modulo** n.*

Fix an integer n. Given two integers a and b, we write $[a]_n = [b]_n$ if the remainders of a and b modulo n are equal, which happens exactly if n divides $a-b$. We alternatively write $a \equiv b \mod n$ in this case, and we say that a ***is congruent to*** b ***modulo*** n. For any integer a, there is exactly one element in the set

$$\left\{[0]_n, [1]_n, [2]_n, \ldots, [n-2]_n, [n-1]_n\right\}$$

which is equal to $[a]_n$. For instance $[3]_4 = [7]_4 = [-1]_4$, and so on. We may perform ***addition of integers modulo*** n, by setting

$$[a]_n + [b]_n := [a+b]_n$$

for any two integers a and b. For instance $[2]_4 + [1]_4 = [3]_4$ and $[2]_5 + [3]_5 = [0]_5$.

One may check that the group properties are transferred from $\mathbb{Z}$ to $\mathbb{Z}_n$:

Proposition 1.27. *The set* $\mathbb{Z}_n := \left\{[0]_n, [1]_n, [2]_n, \ldots, [n-2]_n, [n-1]_n\right\}$ *equipped with addition modulo n is a group.*

Once we define quotients, the groups $\mathbb{Z}_n$ will turn out to be quotients of $\mathbb{Z}$. More on this is in Subsection 2.2.5.

Notation 1.28. If n is understood from the context, we often write elements in $\mathbb{Z}_n$ by simply using a bar instead of square brackets, in order to lighten the notation. For instance in $\mathbb{Z}_4$ we may write

$$\overline{1} + \overline{3} = \overline{0} \quad \text{instead of} \quad [1]_4 + [3]_4 = [0]_4.$$

Example 1.29. The Cayley tables of $\mathbb{Z}_2$, $\mathbb{Z}_3$ and $\mathbb{Z}_4$, with respect to the corresponding addition, are

$+$	$\overline{0}$	$\overline{1}$
$\overline{0}$	$\overline{0}$	$\overline{1}$
$\overline{1}$	$\overline{1}$	$\overline{0}$

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{0}$	$\overline{1}$

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{0}$	$\overline{1}$	$\overline{2}$.

We also define ***multiplication modulo*** n, by setting

$$[a]_n \cdot [b]_n := [a \cdot b]_n$$

for any two integers a and b. For instance $[2]_4 \cdot [1]_4 = [2]_4$ and $[2]_5 \cdot [3]_5 = [1]_5$.

The set $\mathbb{Z}_n$ equipped with multiplication modulo n is *not* a group: associativity and the existence of a neutral element, which is $[1]_n$, are inherited from $\mathbb{Z}$, but for instance $[0]_n$ is not invertible with respect to multiplication. Depending on n, there might be other elements in $\mathbb{Z}_n$ that do not have a multiplicative inverse.

Example 1.30. The Cayley tables of $\mathbb{Z}_2$, $\mathbb{Z}_3$ and $\mathbb{Z}_4$, with respect to the corresponding multiplication, are

$\cdot$	$\overline{0}$	$\overline{1}$
$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{1}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{0}$	$\overline{2}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{2}$	$\overline{1}$.

Observe that in $\mathbb{Z}_4$ there is a non-zero element that is not invertible: $\overline{2}$.

Recall that $\gcd(a,b)$ denotes the greatest common divisor of a and b.

Proposition 1.31. *An element $[x]_n \in \mathbb{Z}_n = \{[0]_n, [1]_n, [2]_n, \dots, [n-2]_n, [n-1]_n\}$ is invertible with respect to multiplication modulo n exactly if $\gcd(x,n) = 1$.*

For instance, in $\mathbb{Z}_{24}$ the element $\overline{1}8$ is not invertible, because $\gcd(18,24) = 6$. On the other hand, the element $\overline{5}$ is invertible. (Indeed, $\overline{5}$ is the inverse of itself, as $\overline{5} \cdot \overline{5} = \overline{1}$ in $\mathbb{Z}_{24}$.)

Proposition 1.32. *The set*

$$\mathbb{Z}_n^\times := \Big\{[k]_n \mid k \in \{0, \dots, n-1\} \text{ and } \gcd(k,n) = 1\Big\}$$

equipped with multiplication modulo n is a group.

Proof. This follows from Lemma 1.25. □

Example 1.33. For every prime $p \in \mathbb{Z}$, we have $\mathbb{Z}_p^\times = \{[1]_p, \dots, [p-1]_p\}$.

The symmetric group S_n, a.k.a. the group of permutations of n elements

Recall from the second item in Examples 1.4 that, given a set X, the usual composition of functions on the set $X^X := \{\text{functions } X \to X\}$ is associative. Moreover, it has a neutral element, consisting of the identity function. Hence, by Lemma 1.25, the following important construction produces a group:

Definition 1.34. Given a set X, denote by

$$S_X := \{\text{invertible functions } X \to X\}.$$

In the special case of $X = \{1, \dots, n\}$, we denote this by

$$S_n := S_{\{1,\dots,n\}}$$

and we call it the ***n-th symmetric group***, or the ***group of permutations of n elements***.[3]

This construction is extremely important: we have not introduced the terminology of subgroups yet, but Cayley's theorem (Theorem 3.2) states that every group is a subgroup of a symmetric group. So, in some sense, understanding symmetric groups is enough to understand all groups. More on this topic is in Section 3.1.

Notation 1.35. A bijection from a finite set to itself, so in particular an element of S_n, is called a ***permutation***. An element $\sigma \in S_n$ is often displayed as

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

Moreover, instead of writing $\sigma \circ \tau$ we usually simply write $\sigma\tau$, for $\sigma, \tau \in S_n$.

Example 1.36. In S_3 we write the permutations

$$\begin{aligned} \sigma(1) &= 1 & \tau(1) &= 3 & \sigma\tau(1) &= 2 \\ \sigma(2) &= 3 & \tau(2) &= 2 & \sigma\tau(2) &= 3 \\ \sigma(3) &= 2 & \tau(3) &= 1 & \sigma\tau(3) &= 1 \end{aligned}$$

as

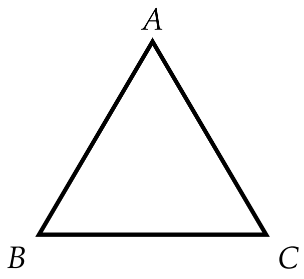
$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \qquad \tau = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \qquad \sigma\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Remark 1.37. For $n \geq 3$, the symmetric group S_n is not abelian. Exercise.

[3] Some authors use a fraktur capital S (the same as for the Sisu candy brand).

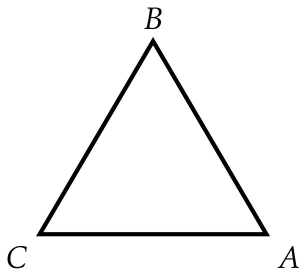
The dihedral group D_n, a.k.a. the group of symmetries of a regular n-gon

Consider an equilateral triangle whose vertices are labeled by A, B and C:



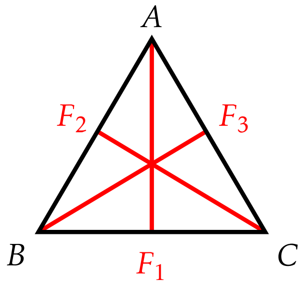
There are two sets of obvious symmetries:

1. One can rotate the triangle through 120° or 240°. Let us denote the clockwise rotation through 120° by R:

   Then the clockwise rotation through 240° can be represented by R^2. If we apply R three times, represented by R^3, we get back where we started. We denote the trivial symmetry by I.

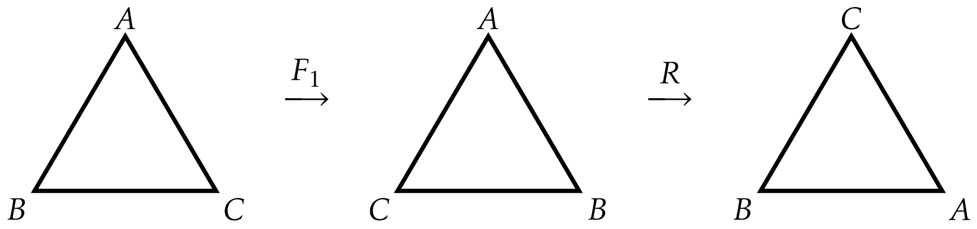
2. The other obvious symmetries are flips. For example, one can draw a vertical line through the top corner and flip about this line. Call this flip F_1. We note that $F_1^2 = I$. There are two other axes to flip about, and we denote the corresponding flips by F_2 and F_3:

The set of symmetries so far is $\{I, R, R^2, F_1, F_2, F_3\}$. This is all, since every symmetry is determined by its action on the vertices of the triangle. There are at most six different permutations of the letters A, B and C. We have given six different symmetries, so we must have given all.

Definition 1.38. The symmetries described above form a group called the ***dihedral group of order*** 3, where the operation is composition, if we see these symmetries as functions.

Remark 1.39. This group is *not abelian*. Consider for instance $R \circ F_1$, which means applying first F_1 and then R:

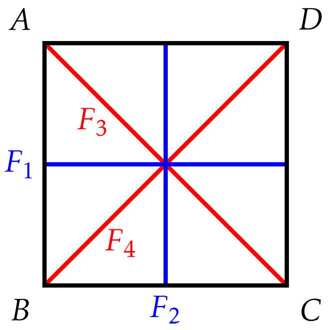


Then $R \circ F_1 = F_3$, since they act in the same way on the vertices. On the other hand, one can see that $F_1 \circ R = F_2$.

For a square, the set of symmetries is similar:

1. One can rotate the square through 90°, call this rotation R. The other roations are obtained as powers of R with itself.

2. There are now four possible flips:



One may in fact construct a similar group for any regular polygon:

Definition 1.40. Given an integer $n \geq 3$, the group of symmetries of a regular n-gon is called the ***dihedral group of order*** n, and it is denoted by D_n.

1.1.5 Powers and order of an element

Let $a, b, c \in G$. An expression of the form $a \cdot b \cdot c$ does not immediately make sense, because a binary operation combines two but not three elements. However, we can consider $(a \cdot b) \cdot c$ and $a \cdot (b \cdot c)$ and by associativity these two expressions are equal. Hence there is no ambiguity in writing $a \cdot b \cdot c$.

Definition 1.41. Let $(G, \cdot)$ be a group and let $a \in G$. For $n \in \mathbb{N}$, define

$$a^n := \underbrace{a \cdot a \cdot \ldots \cdot a}_{n \text{ terms}}, \qquad a^0 := e, \qquad \text{and} \qquad a^{-n} := \underbrace{a^{-1} \cdot a^{-1} \cdot \ldots \cdot a^{-1}}_{n \text{ terms}}.$$

In additive notation, we write

$$na := \underbrace{a + a + \cdots + a}_{n \text{ terms}}, \qquad 0_{\mathbb{Z}}a := 0_G, \qquad (-n)a := \underbrace{(-a) + (-a) + \ldots + (-a)}_{n \text{ terms}}.$$

Remark 1.42 (properties of powers)**.** It can be easily verified that for $n, m \in \mathbb{Z}$, we have

$$a^n \cdot a^m = a^{n+m}, \qquad (a^n)^m = a^{nm} \qquad \text{and} \qquad a^n \cdot a^m = a^m \cdot a^n.$$

Moreover, given a and b such that $ab = ba$, we have $(ab)^n = a^n b^n$. In additive notation, these properties are written

$$na +_G ma = (n +_{\mathbb{Z}} m)a, \qquad m(na) = (mn)a, \qquad na + ma = ma + na.$$

Moreover, if a and b are such that $a + b = b + a$, then $n(a + b) = na + nb$.

Definition 1.43. Let G be a group and let $x \in G$. The ***order of*** x is[4]

$$\pi(x) := \begin{cases} \min\{n \in \mathbb{Z}_{>0} \mid x^n = e\} & \text{if this minimum exists,} \\ +\infty & \text{otherwise.} \end{cases}$$

Examples 1.44.

1. Take $(\mathbb{C}^*, \cdot)$. Then $\pi(i) = 4$ and $\pi(2) = +\infty$.
2. Consider $(\mathbb{Z}_4, +)$. Then $\pi(\overline{1}) = 4$ and $\pi(\overline{2}) = 2$. In general, in $\mathbb{Z}_n$ we have $\pi(\overline{1}) = n$.
3. The elements of the dihedral group D_3 have the following orders:
$$\pi(I) = 1, \qquad \pi(R) = \pi(R^2) = 3, \qquad \pi(F_1) = \pi(F_2) = \pi(F_3) = 2.$$
4. The dihedral group D_n contains a rotation R through $(\frac{360}{n})^\circ$. Then $\pi(R) = n$.
5. The only $a \in G$ with $\pi(a) = 1$ is the identity, $a = e$.
6. The only element of finite order in $(\mathbb{Z}, +)$ is 0.
7. The only elements of finite order in $(\mathbb{R}^*, \cdot)$ are 1 and -1. The order of -1 is 2. If $|x| > 1$, then the sequence x^{2n} is increasing to infinity. If $x^n = 1$ for some positive integer n, then this sequence would be periodic, so $x^n \neq 1$ for every positive integer n. Similar argumentation when $|x| < 1$.
8. Take $\mathrm{GL}_2(\mathbb{R})$ with usual matrix multiplication. The matrix
$$M := \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$
can be thought of as the matrix associated to the plane reflection about the vertical axis. It is the inverse of itself, and $\pi(M) = 2$.

In the following, we use a vertical bar to denote divisibility: for two integers a and b, we write $a|b$, read "a divides b", if there exists an integer q such that $b = qa$.

Proposition 1.45. *Let G be a group and let $x \in G$. For $n, m \in \mathbb{Z}$, we have*

$$x^n = x^m \qquad \Leftrightarrow \qquad \begin{cases} \pi(x) | (n - m) & \text{if } \pi(x) < +\infty \\ n = m & \text{if } \pi(x) = +\infty. \end{cases}$$

Proof. Assume first that $\pi(x) < +\infty$.

[4] Many authors write $|x|$ instead of $\pi(x)$ for the order of x. The notation $\pi(x)$ in this course is meant to avoid confusion with other vertical bars, especially in case you are taking Metric Spaces at the same time, and it comes from the word "period".

($\Rightarrow$) Write $n - m = q\pi(x) + r$, for some q and for some r with $0 \le r < \pi(x)$. Since $x^n = x^m$, we know that $x^n \cdot x^{-m} = e$, hence

$$e = x^n \cdot x^{-m} = x^{n-m} = x^{q\pi(x)+r} = (x^{\pi(x)})^q \cdot x^r = x^r.$$

But $\pi(x)$ is the smallest positive power of x equal to e, and $r < \pi(x)$, and therefore $r = 0$, so that $n - m = q\pi(x)$.

($\Leftarrow$) Let $n - m = k\pi(x)$. Then

$$x^n \cdot x^{-m} = x^{n-m} = x^{k\pi(x)} = (x^{\pi(x)})^k = e^k = e,$$

which means that $x^n = x^m$.

If instead $\pi(x) = +\infty$:

($\Rightarrow$) From $x^n = x^m$ we know that $x^{n-m} = e$, and because the order is infinite this means that $n - m = 0$, that is, $n = m$.

($\Leftarrow$) This is clear.

□

Example 1.46. Take $\mathbb{Z}_{10}$ with multiplication. The group of invertible elements is

$$\mathcal{U}(\mathbb{Z}_{10}) = \{\overline{n} \mid \gcd(n, 10) = 1\} = \{\overline{1}, \overline{3}, \overline{7}, \overline{9}\}.$$

Since $\pi(\overline{3}) = 4$ and 4 divides 12, the proposition above implies for instance that $\overline{3}^{101} = \overline{3}^{113}$.

Corollary 1.47. *If G is a finite group, then every element has finite order.*

Proof. The claim follows immediately from the fact that

$$\#\{x^n \mid n \in \mathbb{Z}\} = \pi(x),$$

so we shall prove this. Indeed, by Proposition 1.45, in case $\pi(x)$ is infinite, we have $x^n \neq x^m$ for $n \neq m$, and in case $\pi(x)$ is finite, we have $\{x^n \mid n \in \mathbb{Z}\} = \{x^0, x^1, \ldots, x^{\pi(x)-1}\}$. □

Example 1.48. The fact that every element has finite order does *not* imply that the group is finite. Consider for instance $(\mathbb{Z}_2[x], +)$, which is infinite. For a polynomial $f \in \mathbb{Z}_2[x]$, we have

$$\pi(f) = \begin{cases} 2 & \text{if } f \neq 0 \\ 1 & \text{if } f = 0. \end{cases}$$

Remark 1.49. The *order of a group G* is the number of elements of G. Since we want to avoid possible confusion with the order of an element, in this course we mostly refer to the "cardinality of G", which we denote $\#G$.

1.2 Subgroups

Given a group $(G, \cdot_G)$, when does it happen that a subset $H \subseteq G$ is itself a group, under the operation of $\cdot_G$ restricted to H?

Definition 1.50. Let $(G, \cdot_G)$ be a group with identity element 1_G. A subset $H \subseteq G$ is called a ***subgroup of*** G if the following conditions hold:

(1) for all $a, b \in H$, we have $a \cdot_G b \in H$;

(2) $1_G \in H$;

(3) for all $h \in H$, the inverse h^{-1} with respect to $\cdot_G$ is in H.

Remark 1.51. 1. The conditions in the definition of a subgroup $H \subseteq G$ mean that, when we restrict the domain of the operation $\cdot_G \colon G \times G \to G$ to $H \times H$, this makes H into a group, meaning that the values are themselves in H, and also the inverses and the identity are in H. There is one more assumption in the definition of a group: associativity of the operation. But this is automatic: if $\cdot_G$ is associative on G, it is still associative on any subset.

2. Note that the second condition does not follow automatically from the other two. Indeed, the empty set satisfies conditions (1) and (3), but $1_G \notin \emptyset$. The point of condition (2) is exactly to make sure that $H \neq \emptyset$. Once we know that $H \neq \emptyset$, then for sure there is some $h \in H$, and by conditions (3) and (1), respectively, we know that $h^{-1} \in H$ and $1_G = h \cdot h^{-1} \in H$. So, we may replace condition (2) in the definition above by "$H \neq \emptyset$".

Examples 1.52. 1. The group G itself is the *improper subgroup* of G. All other subgroups of G are called *proper subgroups*.

2. The singleton $\{1_G\}$ is the *trivial subgroup* of G. All other subgroups of G are called *nontrivial subgroups*.

3. $2\mathbb{Z} := \{2n \mid n \in \mathbb{Z}\}$ is a subgroup of $\mathbb{Z}$ under addition.

4. $\mathbb{R}_{>0} := \{r \in \mathbb{R} \mid r > 0\}$ is a subgroup of $\mathbb{R}^\times := \mathbb{R} \setminus \{0\}$ under multiplication.

5. $H := \{e, R, R^2\}$ is the subgroup of D_3 consisting of the rotations only.

6. The group D_n is a subgroup of $\mathrm{GL}_2(\mathbb{R})$.

7. The set $\{M \in \mathrm{GL}_n(\mathbb{R}) \mid \det(M) > 0\}$ is a subgroup of $\mathrm{GL}_n(\mathbb{R})$.

8. The set $\{M \in \mathrm{GL}_n(\mathbb{R}) \mid \det(M) = \pm 1\}$ is a subgroup of $\mathrm{GL}_n(\mathbb{R})$.

9. The set $\{M \in \mathrm{GL}_n(\mathbb{R}) \mid \det(M) = 1\}$ is a subgroup of $\mathrm{GL}_n(\mathbb{R})$, called the ***special linear group***. It is the intersection of the subgroups in the previous two items in this list.

10. Consider the group $F :=$ {functions $\mathbb{R} \to \mathbb{R}$} under addition. The subset C of F consisting of continuous functions is a subgroup of F.

11. Continuing the previous item, the set $\mathcal{D}$ of differentiable functions $\mathbb{R} \to \mathbb{R}$ is a subgroup of C, and hence a subgroup of F. More generally, for each n we may define the group $\mathcal{D}^n$ of functions from $\mathbb{R}$ to $\mathbb{R}$ which are differentiable n times, and this yields an infinite chain of subgroups

$$\mathcal{D}^0 \quad \supsetneq \quad \mathcal{D}^1 \quad \supsetneq \quad \mathcal{D}^2 \quad \supsetneq \quad \dots$$

12. Polynomials in one variable with real coefficients form a group P, and P is a subgroup of each of the groups $\mathcal{D}^n$ defined above.

Example 1.53. 1. The set $\{\overline{0}, \overline{3}\}$ is not a subgroup of $\mathbb{Z}_4$.

2. The set of odd numbers is not a subgroup of $\mathbb{Z}$.

Proposition 1.54 (Subgroup criterion). *A subset H of a group G is a subgroup if and only if the following hold:*

(i) $H \neq \emptyset$;

(ii) for all $a, b \in H$, *we have* $a \cdot b^{-1} \in H$.

Proof. *Only if:* Since 1_H by condition (2) in the definition, we have $H \neq \emptyset$. Next, let $a, b \in H$. By condition (3), $b^{-1} \in H$, and by condition (1) we have $ab^{-1} \in H$.

If: By (i), we know that there exists some $x \in H$. By (ii), taking $a = b = x$, we get that $xx^{-1} \in H$, but $xx^{-1} = 1_G$, so that we have (2). Given $h \in H$, and since we just learned that $1_G \in H$, by (ii) we get that $h^{-1} = 1_G h^{-1} \in H$, which is (3). Lastly, given $a, b \in H$, we know that $b^{-1} \in H$, so by (ii) we get $ab = a(b^{-1})^{-1} \in H$, which is (1). □

1.2.1 Generating sets

Proposition 1.55 (Intersection of subgroups is a subgroup). *Let G be a group and $\{H_j \mid j \in J\}$ a family of subgroups of G, for some index set J. Then*

$$H := \bigcap_{j \in J} H_j = \{a \in G \mid \textit{for all } j \in J,\ a \in H_j\}$$

is a subgroup of G.

Proof. We will show that the two conditions of the subgroup criterion hold. Since the identity e belongs to every subgroup of G, we have $e \in H_j$ for all $j \in J$, and hence $e \in H$. Let now $a, b \in H$. Then $a, b \in H_j$ for all $j \in J$. Since all the H_j's are subgroups of G, we have $ab^{-1} \in H_j$ for all $j \in J$. Hence $ab^{-1} \in H$. □

Let G be a group and let $S \subseteq G$ be any subset. There is at least one subgroup of G that contains S, namely G itself. By Proposition 1.55, the intersection of all subgroups that contain S is again a subgroup. This is the smallest subgroup of G containing S.

Definition 1.56. Let G be a group and let S be a subset of G. We call ***subgroup generated by*** S, denoted $\langle S \rangle$, the smallest subgroup of G containing S, namely the intersection of all subgroups of G that contain S. The elements of S are called the ***generators*** of $\langle S \rangle$. If S is finite, then $\langle S \rangle$ is said to be ***finitely generated***.

Notation 1.57. If $S = \{g_1, \dots, g_m\}$ is finite, we simply write $\langle g_1, \dots, g_m \rangle$ instead of $\langle \{g_1, \dots, g_m\} \rangle$. In particular, if $S = \{g\}$ is a singleton, we write $\langle g \rangle$ instead of $\langle \{g\} \rangle$.

Remarks 1.58. 1. If $H \subseteq G$ is a subgroup of G, then $\langle H \rangle = H$.

2. The union of subgroups is **not** a subgroup in general. Give examples.

3. The subgroup $\langle e \rangle$ generated by the identity e is the singleton $\{e\}$. Indeed the set $\{e\}$ forms a subgroup, and it is the smallest subgroup containing e.

4. The subgroup $\langle \emptyset \rangle$ generated by the empty set is also equal to the singleton $\{e\}$, since the identity has to be in every subgroup, and $\{e\}$ is therefore the smallest subgroup containing the empty set.

Example 1.59. We have $D_3 = \langle \{R, F_1\} \rangle$.

Examples 1.60. In the following examples, consider $(\mathbb{Z}, +)$ throughout.

1. The subgroup $\langle 0\rangle$ is equal to the singleton $\{0\}$, by the third remark above, since 0 is the identity of $\mathbb{Z}$.

2. The subgroup $\langle 1\rangle$ has to contain 1, and by the definition of subgroup, we also need the sum of any two elements of $\langle 1\rangle$ to still be in $\langle 1\rangle$. So, since $1 \in \langle 1\rangle$, we must have $2 = 1+1 \in \langle 1\rangle$, and therefore also $3 = 2+1 \in \langle 1\rangle$, and similarly for any $n \in \mathbb{N}$ we have $n \in \langle 1\rangle$. And again by definition of subgroup, the inverses of all elements of $\langle 1\rangle$ have to be in $\langle 1\rangle$, but then this means that
$$\langle 1\rangle = \mathbb{Z}.$$
We then say that $\mathbb{Z}$ *is generated by* 1. Similarly, one may see that $\langle -1\rangle = \mathbb{Z}$.

3. The subgroup $\langle 2,3\rangle$ is the smallest subgroup of $\mathbb{Z}$ containing both 2 and 3. This group has to contain $1 = (-2)+3$. But then $\langle 2,3\rangle$ has to contain all of $\mathbb{Z}$ by the same argument as above, and therefore $\langle 2,3\rangle = \mathbb{Z}$.

4. The subgroup $\langle 2\rangle$ is equal to $2\mathbb{Z} = \{\text{even integers}\} = \{2n \mid n \in \mathbb{Z}\}$. Similarly, one has $\langle 3\rangle = \{3n \mid n \in \mathbb{Z}\}$. Exercise.

In particular, $\mathbb{Z}$ is generated by $\{2,3\}$, but it is not generated by either 2 or 3 taken singularly.

1.2.2 Cyclic (sub)groups

Definition 1.61. Let G be a subgroup, a subgroup of G generated by a single element is called a ***cyclic subgroup***. If there exists some $a \in G$ such that $\langle a\rangle = G$, then G is called a ***cyclic group***, and we say that a is a ***generator for*** G.

The previous subsection contains several examples of cyclic (sub)groups. In particular, $(\mathbb{Z},+)$ is cyclic, and it is generated by 1, or by -1.

Proposition 1.62 (Explicit description of cyclic subgroups). *Let G be a group, and let $a \in G$. Then*
$$\langle a\rangle = \{a^i \mid i \in \mathbb{Z}\}.$$

Proof. Set $H := \{a^i \mid i \in \mathbb{Z}\}$. We need to show that H is a subgroup of G, and that H is the smallest subgroup of G that contains a.

We observe that H is a subgroup of G, since it is nonempty and for $a^i, a^j \in H$, we have $a^i(a^j)^{-1} = a^{i-j} \in H$. It is left to show that H is the smallest subgroup containing a. Let K be any subgroup of G containing a. We will show by induction that $a^i \in K$ for $i \in \mathbb{N}$. The base case is $a \in K$. Assume now that $a^n \in K$. Then $a, a^n \in K$ imply $a^{n+1} = aa^n \in K$. We have $e \in K$ by definition of subgroup, and $a^{-i} \in K$ for $i \in \mathbb{N}$ because $(a^i)^{-1} = a^{-i}$. Hence $H \subseteq K$. Since K was any subgroup of G containing a, then H is the smallest subgroup of G containing a. □

Corollary 1.63. *For any $a \in G$, we have $\langle a\rangle = \langle a^{-1}\rangle$.*

Example 1.64. 1. The group $\mathbb{Z}_4$ is generated both by $\overline{1}$ and $\overline{3}$.

2. Consider the symmetric group S_4 and the permutation
$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}.$$
Then we have
$$\sigma^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \qquad \sigma^3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} \qquad \sigma^4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} = \mathrm{id}_{\{1,\dots,4\}},$$
and $\langle \sigma\rangle = \{\mathrm{id}_{\{1,\dots,4\}}, \sigma, \sigma^2, \sigma^3\}$.

3. Consider the dihedral group D_3 and any flip F_i. Then $F_i^2 = I$, and $\langle F_i \rangle = \{I, F_i\}$.

Proposition 1.65. *Let G be a group and let $a \in G$.*

1. *If $\pi(a) = m$, then $\langle a \rangle = \{a^k \mid 0 \leq k \leq m-1\}$ and $\#\langle a \rangle = m$.*

2. *If $\pi(a) = +\infty$, then the map $\mathbb{Z} \to \langle a \rangle$ defined by $n \mapsto a^n$ is bijective.*

Proof. (1) Set $H := \{a^k \mid 0 \leq k \leq m-1\}$. Since $H \subseteq \langle a \rangle$, we only need to prove that $\langle a \rangle \subseteq H$ and $\#\langle a \rangle = m$. This follows from Proposition 1.45.

(2) Suppose a has infinite order. By Proposition 1.62, the mapping $\mathbb{Z} \to \langle a \rangle$ defined by $n \mapsto a^n$ is surjective. It suffices to prove that it is also injective. This also follows from Proposition 1.45. □

Proposition 1.66. *If G is a cyclic group, then G is abelian.*

Proof. Let a be a generator of G. For $x, y \in G$, by Proposition 1.62, there exist $i, j \in \mathbb{Z}$ such that $x = a^i$ and $y = a^j$. It follows that

$$xy = a^i a^j = a^{i+j} = a^{j+i} = a^j a^i = yx.$$

□

Proposition 1.67. *Every subgroup of a cyclic group is a cyclic subgroup.*

Proof. Let G be a cyclic subgroup, and let a be a generator for G, so that $G = \{a^i \mid i \in \mathbb{Z}\}$, by Proposition 1.62. Let H be any subgroup, and denote

$$m := \min\{i \in \mathbb{Z}_{>0} \mid a^i \in H\}.$$

If this minimum does not exist, then the only power of a inside H is a^0, which is the identity of G, so that in this case H is a cyclic subgroup and we are done. So from now on assume that $m > 0$.

We will show that $H = \langle a^m \rangle$. Since $a^m \in H$, the inclusion $\supseteq$ holds, and we need to show $\subseteq$. To this end, let $x \in H$. Because G is generated by a, we know that $x = a^n$, for some n. In order to show that $a^n \in \langle a^m \rangle$, we need to show that n is a multiple of m. We may write $n = qm + r$, with $q, r \in \mathbb{Z}$ and $0 \leq r < m$. Since $a^m \in H$, we have $(a^m)^q = a^{mq} \in H$, and this together with $a^n \in H$, implies, by the subgroup criterion, that

$$a^r = a^{n-qm} = a^n \cdot a^{-qm} = a^n \cdot (a^{qm})^{-1} \in H.$$

But since $r < m$ and a^m is the smallest positive power of a in H, this means that $r = 0$, and then $n = qm$, as we needed to show. □

Corollary 1.68. *The subgroups of $(\mathbb{Z}, +)$ are all of the form $n\mathbb{Z}$, for $n \in \mathbb{N}$.*

Proof. We saw in Example 1.60 that $\mathbb{Z}$ is cyclic, generated by 1. By Proposition 1.67, the subgroups of $\mathbb{Z}$ must be cyclic. If we take $n = 0$, we get the subgroup $\{0\}$. And the positive powers of $1 \in \mathbb{Z}$ are exactly the positive integers. □

1.3 Group homomorphisms

Definition 1.69. Let $(G,*_G)$ and $(H,*_H)$ be two groups. A ***(group) homomorphism from*** $(G,*_G)$ ***to*** $(H,*_H)$ (or simply written "from G to H", if the operations are understood from the context) is a function $f\colon G \to H$ that satisfies

$$f(a *_G b) = f(a) *_H f(b),$$

for all $a,b \in G$. Homomorphisms satisfiying additional conditions deserve special names:

- An injective homomorphism is called a ***monomorphism***.
- A surjective homomorphism is called an ***epimorphism***.
- A bijective homomorphism is called an ***isomorphism***. An isomorphism from G to the same group G is called an ***automorphism***. The set of all automorphisms of a group is denoted $\mathrm{Aut}(G)$.

It might be helpful to visualize pictorially a homomorphism $f\colon G \to H$ as follows:

$$\begin{tikzcd}
G \times G \arrow[r, "f \times f"] \arrow[d, "*_G"'] & H \times H \arrow[d, "*_H"] \\
G \arrow[r, "f"'] & H
\end{tikzcd}
\qquad
\begin{tikzcd}
(a,b) \arrow[r, mapsto] \arrow[d, mapsto] & (f(a), f(b)) \arrow[d, mapsto] \\
a *_G b \arrow[r, mapsto] & f(a *_G b) = f(a) *_H f(b)
\end{tikzcd}$$

The diagram "commutes", meaning that for any element $(a,b) \in G \times G$ in the top-left corner, either composition leading to the bottom-right corner gives the same result.

Example 1.70.

1. The determinant is a homomorphism from $\mathrm{GL}_n(\mathbb{R})$ to $\mathbb{R}^*$, since $\det(AB) = \det(A)\det(B)$. It is in fact an epimorphism.
2. Derivative is a homomorphism, since $(f+g)' = f'+g'$. From what group to what group?
3. For any integer n, the map $\phi_n\colon \mathbb{Z} \to \mathbb{Z}$ defined by $\phi_n(m) := nm$ is a homomorphism. The map ϕ_0 is the trivial homomorphism, ϕ_1 is the identity map and ϕ_{-1} maps $\mathbb{Z}$ onto $\mathbb{Z}$. In all other cases, the map ϕ_n is not onto $\mathbb{Z}$.
4. For any integer n, the map $p_n\colon \mathbb{Z} \to \mathbb{Z}_n$ given by $p_n(m) := [m]_n$ for all $m \in \mathbb{Z}$ is an epimorphism.
5. Fix a matrix $A \in M_{n,m}(\mathbb{R})$. From some matrix algebra course, you know that the map

   $$\begin{aligned} f_A\colon \mathbb{R}^m &\longrightarrow \mathbb{R}^n \\ v &\longmapsto Av. \end{aligned}$$

   is a group homomorphism: for all $v,w \in \mathbb{R}^m$, we have $f_A(v+w) = f_A(v) + f_A(w)$
6. More generally, if V and W are vector spaces, then V and W are abelian groups under addition. If $T\colon V \to W$ is a linear transformation, then in particular $T(u+v) = T(u) + T(v)$. Hence T is a homomorphism from V to W. We simply forget the fact that linear transformations are also well-behaved with respect to scalar multiplication.
7. For any group G, the identity map $\mathrm{id}_G\colon G \to G$ is an isomorphism. In general, for any subgroup $H \subseteq G$, the inclusion map $H \hookrightarrow G$ is a homomorphism.

Proposition 1.71. *Let G and H be groups and $f\colon G \to H$ a group homomorphism. Then the following statements hold:*

1. *If e_G is the identity of G and e_H the identity of H, then $f(e_G) = e_H$.*
2. *We have $f(a^{-1}) = f(a)^{-1}$ for all $a \in G$.*
3. *If K is a subgroup of G , then $f(K)$ is a subgroup of H.*
4. *If L is a subgroup of H, then $f^{-1}(L) := \{a \in G \mid f(a) \in L\}$ is a subgroup of G.*
5. *If a has finite order in G, then $f(a)$ has finite order in H.*
6. *If L is a group and $g\colon H \to L$ is a homomorphism, then $g \circ f\colon G \to L$ is a homomorphism.*
7. *Assume $f\colon G \to H$ is surjective (i.e., an epimorphism). If G is abelian, then H is abelian.*

Proof. 1. We have

$$f(e_G) = f(e_G e_G) = f(e_G)f(e_G),$$

where the last equality holds since f is a homomorphism. By multiplying both sides by $f(e_G)^{-1}$, we get $e_H = f(e_G)$.

2. We have

$$f(a)f(a^{-1}) = f(aa^{-1}) = f(e_G) = e_H,$$

and similarly $f(a^{-1})f(a) = e_H$. Hence $f(a)^{-1} = f(a^{-1})$.

3. We will show that $f(K)$ is a subgroup of H by using the subgroup criterion. The set $f(K)$ is nonempty, since $e_H \in f(K)$. Let $b_1, b_2 \in f(K)$. Then there exist $a_1, a_2 \in K$ such that $f(a_1) = b_1$ and $f(a_2) = b_2$. Then

$$b_1 b_2^{-1} = f(a_1)f(a_2)^{-1} = f(a_1)f(a_2^{-1}) = f(a_1 a_2^{-1}).$$

Since K is a subgroup of G, we have $a_1 a_2^{-1} \in K$ and $b_1 b_2^{-1} = f(a_1 a_2^{-1}) \in f(K)$.

4. We will show that $f^{-1}(L)$ is a subgroup of G by using the subgroup criterion. The set $f^{-1}(L)$ is nonempty, since $e_G \in f^{-1}(L)$. Let $a_1, a_2 \in f^{-1}(L)$. Then $b_1 := f(a_1)$ and $b_2 := f(a_a)$ are elements of L. Thus

$$f(a_1 a_2^{-1}) = f(a_1)f(a_2^{-1}) = f(a_1)f(a_2)^{-1} = b_1 b_2^{-1} \in L,$$

because L is a subgroup of H. Thus $a_1 a_2^{-1} \in f^{-1}(L)$.

5. Since $a \in G$ is of finite order, there exists $n \in \mathbb{N}$ such that $a^n = e_G$. Then

$$f(a)^n = f(a^n) = f(e_G) = e_H.$$

This shows that $f(a)$ has finite order in H.

6. Let $a_1, a_2 \in G$. Then

$$(g \circ f)(a_1 a_2) = g(f(a_1 a_2)) = g(f(a_1)f(a_2)) = g(f(a_1))g(f(a_2)) = (g \circ f)(a_1) \cdot (g \circ f)(a_2)$$

7. Let $a', b' \in H$. There exist $a, b \in G$ such that $f(a) = a'$ and $f(b) = b'$. We have

$$a'b' = f(a)f(b) = f(ab) = f(ba) = f(b)f(a) = b'a',$$

so H is abelian.

□

1.3.1 Isomorphisms

Proposition 1.72. *The composition of isomorphisms is an isomorphism, and the inverse of an isomorphism is an isomorphism. More precisely:*

(1) Let $f\colon G \to H$ and $g\colon H \to K$ be isomorphisms. Then $g \circ f\colon G \to K$ is an isomorphism.

(2) Let $f\colon G \to H$ be an isomorphism. Then $f^{-1} : H \to G$ is an isomorphism.

Proof. (1) We saw in the previous proposition that the composition of homomorphisms is a homomorphism. Moreover, the composition of bijections is a bijection.

(2) Let $b_1, b_2 \in H$. Then

$$f(f^{-1}(b_1)f^{-1}(b_2)) = f(f^{-1}(b_1))f(f^{-1}(b_2)) = b_1 b_2.$$

Hence

$$f^{-1}(b_1)f^{-1}(b_2) = f^{-1}(b_1 b_2).$$

Moreover, the inverse of a bijection is a bijection. □

Corollary 1.73. *Let G be a group. The set* $\mathrm{Aut}(G)$ *is a group with respect to composition.*

Proof. Associativity of compositions was an exercise. By the previous proposition, the set $\mathrm{Aut}(G)$ is closed under taking compositions, and the inverse of an element of $\mathrm{Aut}(G)$ is in $\mathrm{Aut}(G)$. Lastly, the identity map id_G is in $\mathrm{Aut}(G)$. □

The proposition and corollary above also follow from Lemma 1.25, since composition is an associative operation on the set $\mathrm{Hom}(G,G) := \{\text{homomorphisms } G \to G\}$, and the automorphisms of G are exactly the invertible elements under this operation.

Examples 1.74.

1. For any group G, the identity map

$$\begin{aligned} \mathrm{id}_G\colon G &\longrightarrow G \\ x &\longmapsto x \end{aligned}$$

is an automorphism of G. This is the *trivial automorphism* of G.

2. The map

$$\begin{aligned} \mathbb{Z} &\longrightarrow \mathbb{Z} \\ n &\longmapsto -n \end{aligned}$$

is the only non-trivial automorphism of $\mathbb{Z}$.

3. Consider the set $\{a,b\}$, with $a \neq b$, and define on this set the operation

$*$	a	b
a	a	b
b	b	$a.$

Then $(\{a,b\},*)$ is a group, and it is isomorphic to $(\mathbb{Z}_2,+)$.

4. The logarithm from $\mathbb{R}^+$ to $\mathbb{R}$ is an isomorphism, where we consider $\mathbb{R}^+$ under multiplication and $\mathbb{R}$ under addition. The inverse of the logarithm is the exponential map from $\mathbb{R}$ to $\mathbb{R}^+$. The logarithm is an homomorphism since $\log(xy) = \log(x) + \log(y)$.

5. Let G be a group, and let $a \in G$. The map
$$\begin{aligned} c_a \colon G &\longrightarrow G \\ x &\longmapsto axa^{-1} \end{aligned}$$
is an automorphism of G. This automorphism is called ***conjugation by*** $\boldsymbol{a}$.

6. Any two cyclic groups of the same order are isomorphic.

7. Any cyclic group is isomorphic to either $\mathbb{Z}$ or $\mathbb{Z}_n$, for a suitable n. Exercise.

8. The groups S_3 and D_3 are isomorphic. For $n > 3$, this is not true for obvious reasons of cardinality, but D_n is still isomorphic to a subgroup of S_n. Exercise.

Definition 1.75. If there exists an isomorphism between two groups G and H, we say that G and H are ***isomorphic***, and we denote this by writing $G \cong H$.

As a consequence of Proposition 1.72, we get the following:

Corollary 1.76. *Isomorphism is an equivalence relation.*

Proof. Let G be a group. Then $G \cong G$ because $\mathrm{id}_G \colon G \to G$ is an isomorphism, and this means that $\cong$ is a reflexive relation. Transitivity and symmetry of $\cong$ follow directly from parts (1) and (2) of Proposition 1.72, respectively. □

1.3.2 Kernel and image of a homomorphism

Definition 1.77. Let $f \colon G \to H$ be a group homomorphism. The ***kernel of*** f is
$$\ker(f) := \{g \in G \mid f(g) = e_H\} \quad \subseteq G.$$
The ***image*** of f is
$$\operatorname{im}(f) := \{f(g) \mid g \in G\} \quad \subseteq H.$$

Example 1.78. Fix a matrix $A \in M_{n,m}(\mathbb{R})$. The map
$$\begin{aligned} f_A \colon \mathbb{R}^m &\longrightarrow \mathbb{R}^n \\ v &\longmapsto Av \end{aligned}$$
is a group homomorphism. The kernel of f_A is usually called the *nullspace* of A, and it is the set $\{v \in \mathbb{R}^m \mid Av = \mathbf{0}\}$.

Proposition 1.79. *Let $f \colon G \to H$ be a homomorphism. Then:*

- $\ker(f)$ *is a subgroup of G,*
- $\operatorname{im}(f)$ *is a subgroup of H.*

Proof. The statements are special cases of parts (4) and (3) of Proposition 1.71, respectively. More precisely, $\ker(f) = f^{-1}(\{e_H\})$ is a subgroup of G because $\{e_H\}$ is a subgroup of H, and $\operatorname{im}(f) = f(G)$ is a subgroup of H because G is a subgroup of G. □

In fact we will prove in Proposition 2.36 something stronger for the kernel, namely that the kernel is a *normal* subgroup.

Proposition 1.80 (Criterion for injectivity). *Let $f\colon G \to H$ a homomorphism. Then*

$$f \text{ is injective} \quad \Leftrightarrow \quad \ker(f) = \{e_G\}.$$

Proof. ($\Rightarrow$) Assume f is injective. By Proposition 1.71, we have $f(e_G) = e_H$, namely $e_G \in \ker(f)$. By injectivity, only one element can map to e_H, so that $\ker(f) = \{e_G\}$.

($\Leftarrow$) Suppose now that $\ker(f) = \{e_G\}$. Let $a, b \in G$ be such that $f(a) = f(b)$. Then $e_H = f(a) \cdot f(b)^{-1} = f(a) \cdot f(b^{-1}) = f(ab^{-1})$. Hence $e_G = ab^{-1}$, which means that $a = b$. □

Remark 1.81. You studied in your first matrix algebra course matrices that have a trivial nullspace, namely consisting of the zero vector only. You may want to go back to your notes from that time and compare them with the result above. You will see that if the system $A\mathbf{x} = \mathbf{0}$ has a unique solution (necessarily $\mathbf{x} = \mathbf{0}$), then for any vector $\mathbf{b}$ the more general system $A\mathbf{x} = \mathbf{b}$ also has a unique solution. This is exactly a special case of the proposition above.

Corollary 1.82. *Let $f\colon G \to H$ be a monomorphism (i.e., an injective homomorphism). Then, for all $x \in G$, we have $\pi(x) = \pi(f(x))$.*

Proof. In general, we have $f(x)^n = f(x^n)$. By Proposition 1.80, we have $f(x^n) = e_H$ if and only if $x^n = e_G$. □

Chapter 2

Quotients

The idea of taking a quotient of a set X by an equivalence relation is that you start from X, and you "glue" some elements together, thereby identifying them. In this course we not only do this for sets, but for groups, too. For instance, starting from $X = \mathbb{Z}$ we identify all even numbers with each other, call their set E, and all odd numbers with each other, call their set O. We define a group operation on the set $\{E, O\}$ by setting

$+$	E	O
E	E	O
O	O	E

which amounts to saying that

+	even	odd
even	even plus even is even	even plus odd is odd
odd	odd plus even is odd	odd plus odd is even.

Indeed this yields a group. With the notation introduced in Section 2.2.5, this group is called $\mathbb{Z}/2\mathbb{Z}$, and it is isomorphic to $\mathbb{Z}_2$.

2.1 Relations and equivalence relations

Some familiarity with equivalence relations is assumed for this course. Nevertheless, in this section we recall the definition of that and related notions.

Definition 2.1. Let X be a set. A ***(binary) relation on*** X is a subset $R \subseteq X \times X$. We write xRy as shorthand for $(x, y) \in R$.

Examples 2.2.

1. The set $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x - y \leq 0\}$ is a relation on $\mathbb{R}$.

2. Let $P :=$ {points on Earth}. The set $J := \Big\{(p, q) \in P \times P \mid \text{you can jump from } p \text{ to } q\Big\}$ is a relation on P.

3. Let $n \in \mathbb{Z}_{>0}$. Recall from Subsection 1.1.4 that for two integers a and b, we say that a and b are *congruent modulo n* if n divides $a - b$, and in this case we write $[a]_n = [b]_n$, or alternatively $a \equiv b \mod n$. The set $C := \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid [a]_n = [b]_n\}$, is a relation on $\mathbb{Z}$.

4. Consider $X := \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ and the following relation on X:

$$(a, b)R(c, d) \qquad \text{if} \qquad ad = bc.$$

Formally, R is a subset of $X \times X = \Big(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\Big) \times \Big(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\Big)$.

Definition 2.3. A binary relation R on a set X is called an ***equivalence relation*** if R is:

1. ***reflexive***, i.e., for every $x \in X$, we have xRx,

2. ***symmetric***, i.e., for every $x, y \in X$, if xRy, then yRx,

3. ***transitive***, i.e., for every $x, y, z \in X$, if xRy and yRz, then xRz.

Examples 2.4.

1. For any set X, the set $X \times X$ is the largest relation on X, and it is an equivalence relation on X.

2. For any set X, the set $\{(a,a) \mid a \in X\}$ is the smallest equivalence relation on X.

3. In Examples 2.2, the first two relations are not equivalence relations. Indeed, the first one is not symmetric: if $x - y \leq 0$, it does not hold in general that $y - x \leq 0$. And the relation J is not transitive: it's not true in general that if you can jump from x to y and from y to z, then you can also jump from x to z.

4. We show that the relation C in item 3 of Examples 2.2 is an equivalence relation, for a fixed $n \in \mathbb{Z}_{>0}$:

   - C is reflexive: For any $a \in \mathbb{Z}$, it is true that n divides $a - a = 0$, since $n \cdot 0 = 0$. Hence aCa.
   - C is symmetric: Suppose aCb, that is, n divides $a - b$. This means that $nq = a - b$ for some q, but then n divides $b - a$, since we can write $n(-q) = b - a$, and this means that bCa.
   - C is transitive: Suppose aCb and bCc, which means that we have $nq = a - b$ and $nq' = b - c$. But then $n(q + q') = (a - b) + (b - c) = a - c$, so that n divides $a - c$, which means that aCc.

5. You may check that the relation R in item 4 of Examples 2.2 is an equivalence relation as an exercise.

6. Consider the set $X := \{1, 2, 3\}$ and the relations

$$R_1 := \Big\{(1,1), (1,2), (2,1), (2,2)\Big\} \quad \text{and} \quad R_2 := \Big\{(1,1), (1,2), (2,2), (3,3)\Big\}.$$

   Neither is an equivalence relation: the relation R_1 is not reflexive, because $(3,3) \notin R_1$, and the relation R_2 is not symmetric, because we have $(1,2) \in R_2$ but $(2,1) \notin R_2$.

7. Another example of an equivalence relation is given in Definition 2.14.

Many authors tend to use the symbol $\sim$ instead of R for an equivalence relation.

Definition 2.5. Let $\sim$ be an equivalence relation on a set X. For any element $x \in X$, the set

$$[x]_\sim := \{y \in X \mid x \sim y\}$$

is called the ***equivalence class of*** x. An element of an equivalence class is called a ***representative*** of that equivalence class.[1]

[1] People often write just $[x]$, and in this course we will often write $\overline{x}$ for the sake of brevity, to denote the equivalence class of x.

Example 2.6. Consider the relation in item 3 of Examples 2.2, namely congruence modulo n. The equivalence class of $a \in \mathbb{Z}$ with respect to this relation is

$$[a]_n = \{b \in \mathbb{Z} \mid a \equiv b \mod n\}.$$

For instance, if we pick $n = 2$, we have

$$[24]_2 = \{\text{even integers}\} \qquad \text{and} \qquad [17]_2 = \{\text{odd integers}\},$$

and for instance $[0]_2 = [24]_2 = [56]_2 = [1000]_2$, etc. There are no other equivalence classes. In general, we do not use random-looking representatives like $[24]_2$, but instead it is more canonical to use the representative included between 0 and $n-1$, so that in the case of $n = 2$ the two equivalence classes are $[0]_2$ and $[1]_2$. If instead we pick $n = 3$, there are three equivalence classes, which are $[0]_3$, $[1]_3$ and $[2]_3$. Each integer belongs to exactly one equivalence class.

Definition 2.7. Let $\sim$ be an equivalence relation on a set X. For all x, denote by $\overline{x}$ its equivalence class. The ***quotient of*** X ***by*** $\sim$, denoted $X/\sim$, is the set of equivalence classes of the elements of X:

$$X/\!\sim := \{\overline{x} \mid x \in X\}.$$

Example 2.8. Consider again congruence modulo n. The quotient modulo this relation is

$$\Big\{[0]_n, [1]_n, [2]_n \ldots, [n-1]_n\Big\}.$$

The set $\mathbb{Z}_n$ now has a deeper meaning than just the "set of symbols of the form $[i]_n$".

Example 2.9 (Definition of $\mathbb{Q}$)**.** (*This example might be misleading. For now focus on the set-theoretic level, not on the algebra.*) Consider the equivalence relation $\sim$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ defined by

$$(a,b) \sim (c,d) \qquad \text{if} \qquad ad = bc.$$

(This is the relation R in item 4 of Examples 2.2.) The equivalence class $\overline{(a,b)}$ of $(a,b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is usually denoted by $\frac{a}{b}$. This is a way of constructing the set of rational numbers as a quotient, provided that we have constructed the set $\mathbb{Z}$ first:

$$\mathbb{Q} := \Big(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\Big)/\!\sim .$$

Remark 2.10. You need to be *careful* when dealing with a function whose domain is a quotient. For instance the "function"

$$\begin{aligned} f\colon \mathbb{Z}_n &\longrightarrow \mathbb{Z} \\ [i]_n &\longmapsto i+3 \end{aligned}$$

is *not* a function. It is not "well-defined", the output depends on which representative we choose for the input: for instance we have $[0]_n = [n]_n$, but unfortunately $f([0]_n) = 3$ and $f([n]_n) = n + 3$, and if $n \neq 0$ the two outputs 3 and $n + 3$ are different. Instead, the function

$$\begin{aligned} g\colon \mathbb{Z}_6 &\longrightarrow \mathbb{Z}_3 \\ [i]_6 &\longmapsto [i]_3 \end{aligned}$$

is well-defined! Indeed, take two representatives i and j for the same equivalence class in $\mathbb{Z}_6$, meaning that $[i]_6 = [j]_6$, which more explicitly means that $i - j$ is a multiple of 6, or in other words $i - j = 6q$ for some $q \in \mathbb{Z}$. But then $i - j = 3 \cdot (2q)$, so that $[i]_3 = [j]_3$. Hence, the output of the function does not depend on the choice of a representative for the equivalence class in the input of g. (This worked because 6 is a multiple of 3. If instead you try to define the analogous "function" $\mathbb{Z}_3 \to \mathbb{Z}_6$, that turns out to be again not a function.)

Definition 2.11. Let X be a set. A ***partition of*** X is a family $P = \{X_i \mid i \in I\}$ of subsets of X, where I is some index set, such that the following conditions hold:

1. the X_i's cover X, that is, $\bigcup_{i\in I} X_i = X$,
2. the X_i's are disjoint, that is, if $i \neq j$, then $X_i \cap X_j = \emptyset$,
3. the X_i's are non-empty, that is, $X_i \neq \emptyset$ for all $i \in I$.

Lemma 2.12. *Let X be a set. The map*

$$\begin{aligned}\Phi\colon \{\textit{equivalence relations on } X\} &\longrightarrow \{\textit{partitions of } X\}\\ \sim &\longmapsto X/\sim\end{aligned}$$

is a bijection.

Proof. The fact that for any equivalence relation $\sim$ on X the quotient $\Phi(\sim) = X/\sim$ is indeed a partition of X follows from the definition of equivalence relation. We leave the details and the fact that Φ is a bijection as an exercise. □

Example 2.13. (This example has no algebraic meaning, it is just meant to illustrate the above lemma.) Consider the set $X = \{1,2,3\}$. There are five possible partitions of X, and they are listed below (on the right), next to the corresponding equivalence relation on X (on the left):

$$\begin{aligned}
\sim = \Big\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\Big\} &\quad\leftrightarrow\quad P = \Big\{\{1,2,3\}\Big\}\\
\sim = \Big\{(1,1),(1,2),(2,1),(2,2),(3,3)\Big\} &\quad\leftrightarrow\quad P = \Big\{\{1,2\},\{3\}\Big\}\\
\sim = \Big\{(1,1),(1,3),(2,2),(3,1),(3,3)\Big\} &\quad\leftrightarrow\quad P = \Big\{\{1,3\},\{2\}\Big\}\\
\sim = \Big\{(1,1),(2,2),(2,3),(3,2),(3,3)\Big\} &\quad\leftrightarrow\quad P = \Big\{\{\{1\},\{2,3\}\Big\}\\
\sim = \Big\{(1,1),(2,2),(3,3)\Big\} &\quad\leftrightarrow\quad P = \Big\{\{1\},\{2\},\{3\}\Big\}.
\end{aligned}$$

(Remember that formally a relation is a subset of $X \times X$. The first relation, corresponding to the trivial partition $\{\{1,2,3\}\}$ means that all elements are in relation with each other, and indeed there is just a single equivalence class. The last relation means that each element is only in relation with itself, and indeed there are three equivalence classes.)

2.1.1 The first isomorphism theorem for sets

Definition 2.14. Given a function $f\colon X \to Y$ between sets, we define a relation $\sim_f$ on X as follows:

$$x \sim_f y \qquad \text{if} \qquad f(x) = f(y).$$

Remark 2.15. The relation $\sim_f$ is an equivalence relation.

Proof.

- Reflexivity: For any $x \in X$, we clearly have $f(x) = f(x)$, hence $x \sim_f x$.
- Symmetry: Suppose $x \sim_f y$, that is, $f(x) = f(y)$. Then of course $f(y) = f(x)$, so that $y \sim_f x$.
- Transitivity: Suppose $x \sim_f y$ and $y \sim_f z$, which means that we have $f(x) = f(y)$ and $f(y) = f(z)$. But then $f(x) = f(z)$, so that $x \sim_f z$.

Note that each property follows from the respective property for equality, which indeed is an equivalence relation. □

Theorem 2.16 (First Isomorphism Theorem for Sets). *Let $f\colon X \to Y$ be a function between sets. Consider the equivalence relation $\sim_f$ defined above. For any $x \in X$, write $\overline{x}$ for the equivalence class of x, and define the function $p\colon X \to X/{\sim_f}$ by $p(x) := \overline{x}$. Then there is a unique function φ that makes the following digram commute*

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "p"'] & Y \\
X/{\sim_f} \arrow[ur, dashed, "\varphi"'] &
\end{tikzcd}$$

(i.e., there is a unique function φ such that $\varphi \circ p = f$), and this is defined by $\varphi(\overline{x}) := f(x)$. The function φ is injective, thereby defining a bijection between $X/{\sim_f}$ and the image of f.

Proof. First of all we check that the map φ defined by $\varphi(\overline{x}) := f(x)$ is indeed well-defined, namely that the definition does not depend on the choice of a representative for a given equivalence class. Let x' be a representative of the class of x, meaning that $\overline{x} = \overline{x}'$, or more explicitly $x \sim_f x'$. By definition of $\sim_f$, this means that $f(x) = f(x')$. So indeed φ is well-defined, because then $\varphi(\overline{x}) = \varphi(\overline{x'})$.

Next we check that φ makes the diagram commute: indeed we have

$$\varphi \circ p(x) = \varphi(p(x)) = \varphi(\overline{x}) = f(x).$$

And we check that φ is the *unique* function that makes the diagram commute: Suppose a function k satisfies $k \circ p = f$, which means that for all $x \in X$, we have $k(p(x)) = f(x)$. Since $p(x) = \overline{x}$, this tells us exactly that k has to satisfy $k(\overline{x}) = f(x)$ for all $x \in X$, but then $k = \varphi$.

Now we check that φ is injective: If $\varphi(\overline{x}) = \varphi(\overline{x'})$, by definition of φ we immediately have $f(x) = f(y)$, which means that $x \sim_f y$, or in other words $\overline{x} = \overline{x}'$.

Lastly, we need to show that φ induces a bijection with $\mathrm{im}(f)$, namely we show that $\mathrm{im}(f) = \mathrm{im}(\varphi)$, by proving both inclusions:

($\subseteq$) If $b \in \mathrm{im}(f)$, we have $b = f(a)$ for some $a \in X$, but $f(a) = \varphi \circ p(a) = \varphi(\overline{a})$, so that $b \in \mathrm{im}(\varphi)$.

($\supseteq$) If $b \in \mathrm{im}(\varphi)$, we have $b = \varphi(\overline{a})$ for some $\overline{a} \in X/{\sim_f}$, but $\varphi(\overline{a}) = f(a)$, so that $b \in \mathrm{im}(f)$.

□

Example 2.17. Consider the function

$$f\colon \mathbb{Z} \longrightarrow \{a,b\}, \qquad x \longmapsto \begin{cases} a & \text{if } x \text{ is even,} \\ b & \text{if } x \text{ is odd.} \end{cases}$$

The relation $\sim_f$ splits $\mathbb{Z}$ in two equivalence classes that we are already familiar with:

$$[0]_{\sim_f} = \{\text{even integers}\} \qquad \text{and} \qquad [1]_{\sim_f} = \{\text{odd integers}\}.$$

By the First Isomorphism Theorem, there is a bijection between $\mathbb{Z}/{\sim_f}$ and the image of f, which is the whole set $\{a,b\}$. This is a complicated way of saying that in $\mathbb{Z}$ there are some even numbers, some odd numbers, and nothing else.

2.2 Quotients of groups

In this section we study quotients of groups: given a group G, we study certain equivalence relations $\sim$ whose corresponding set-theoretic quotient $G/{\sim}$ can be given a group structure.

2.2.1 Cosets

Notation 2.18. Given two *subsets* A and B of a group $(G,\cdot)$, we denote

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

In case $A=\{a\}$, we write $aB := \{a\}B$, and similarly if $B=\{b\}$, we write $Ab := A\{b\}$. In additive notation, namely if the operation is written with the symbol $+$, we write

$$A+B := \{a+b \mid a\in A, b\in B\},$$

and so on.

Remarks 2.19. You proved the following properties in the first problem set:

- given three subsets A, B and C of a group G, we have $(AB)C = A(BC)$;
- for a *subgroup* $H\subseteq G$, we have $HH = H$.

Definition 2.20. Let $(G,\cdot)$ be a group. Given a subgroup $H\subseteq G$ and $x\in G$, the subsets of G

$$Hx = \{h\cdot x \mid h\in H\} \qquad \text{and} \qquad xH = \{x\cdot h \mid h\in H\}$$

are called the ***right coset of** H **with respect to** x* and the ***left coset of** H **with respect to** x*, respectively. In additive notation, we write

$$H+x = \{h+x \mid h\in H\} \qquad \text{and} \qquad x+H = \{x+h \mid h\in H\}.$$

Examples 2.21.

- Consider the subgroups $H := \{I,R,R^2\}$ and $K := \{I,F_1\}$ of D_3. The right and left cosets of K are

$$\begin{aligned} KI &= K & IK &= K\\ KR &= \{R, F_2\} & RK &= \{R, F_3\}\\ KR^2 &= \{R^2, F_3\} & R^2K &= \{R^2, F_2\}\\ KF_1 &= \{F_1, I\} & F_1K &= \{F_1, I\}\\ KF_2 &= \{F_2, R\} & F_2K &= \{F_2, R^2\}\\ KF_3 &= \{F_3, R^2\} & F_3K &= \{F_3, R\}. \end{aligned}$$

The right and left cosets of H are

$$\begin{aligned} HI &= H & IH &= H\\ HR &= \{R, R^2, I\} & RH &= \{R, R^2, I\}\\ HR^2 &= \{R^2, I, R\} & R^2H &= \{R^2, I, R\}\\ HF_1 &= \{F_1, F_3, F_2\} & F_1H &= \{F_1, F_2, F_3\}\\ HF_2 &= \{F_2, F_1, F_3\} & F_2H &= \{F_2, F_3, F_1\}\\ HF_3 &= \{F_3, F_2, F_1\} & F_3H &= \{F_3, F_1, F_2\}. \end{aligned}$$

You will probably notice that for each $x\in D_3$, we have $Hx = xH$, although the same does not hold for K. We will see that this is because H is a *normal* subgroup, and K is not.

- Consider the subgroup $2\mathbb{Z}$ of $\mathbb{Z}$. For instance for the element $1 \in \mathbb{Z}$, the right and left coset of $2\mathbb{Z}$ with respect to 1 are

$$\begin{aligned} 2\mathbb{Z}+1 &= \{n+1 \mid n \in 2\mathbb{Z}\} = \{\text{odd integers}\} \\ 1+2\mathbb{Z} &= \{1+n \mid n \in 2\mathbb{Z}\} = \{\text{odd integers}\}, \end{aligned}$$

  and we get the same cosets if we replace 1 by any other odd integer. Instead, the right and left coset of $2\mathbb{Z}$ with respect to 4 are

$$\begin{aligned} 2\mathbb{Z}+4 &= \{n+4 \mid n \in 2\mathbb{Z}\} = \{\text{even integers}\} \\ 4+2\mathbb{Z} &= \{4+n \mid n \in 2\mathbb{Z}\} = \{\text{even integers}\}, \end{aligned}$$

  and we get the same cosets if we replace 4 by any other even integer. Note that in this case left and right cosets with respect to the same element will always be equal, since the group is abelian.

Recall that, given any function f, one may construct the equivalence relation $\sim_f$ as in Definition 2.14. The two equivalence relations defined below are special cases of that construction.

Definition 2.22. For a subgroup $H \subseteq G$, we define two relations: for $x, y \in G$, we write

$$x \sim_H y \quad \text{if} \quad Hx = Hy, \qquad \text{and} \qquad x \approx_H y \quad \text{if} \quad xH = yH.$$

Remarks 2.23.

- The relations $\sim_H$ and $\approx_H$ are indeed special cases of the relation $\sim_f$ in Definition 2.14: the function to consider for $\sim_H$ and $\approx_H$ is respectively

$$\begin{aligned} f\colon G &\longrightarrow \{\text{subsets of } G\} \\ x &\longmapsto Hx, \end{aligned} \qquad\qquad \begin{aligned} g\colon G &\longrightarrow \{\text{subsets of } G\} \\ x &\longmapsto xH. \end{aligned}$$

  Of course $x \sim_f y$ means exactly that $x \sim_H y$, and analogously $x \sim_g y$ means that $x \approx_H y$. Therefore $\sim_H$ and $\approx_H$ are equivalence relations.

- We may rewrite in alternative ways the condition imposed by the equivalence relations just introduced: for $x, y \in G$, we have

$$x \sim_H y \quad \Leftrightarrow \quad y \cdot x^{-1} \in H \quad \Leftrightarrow \quad y \in Hx$$

  and

$$x \approx_H y \quad \Leftrightarrow \quad x^{-1} \cdot y \in H \quad \Leftrightarrow \quad y \in xH.$$

- A consequence of the previous point is that the equivalence class of x with respect to $\sim_H$ is exactly the right coset Hx. And the equivalence class of x with respect to $\approx_H$ is the left coset xH.

Example 2.24. Consider the subgroup $K = \{I, F_1\}$ of D_3. In the first item of Examples 2.21, we computed the right and left cosets of K. The equivalence classes of $\sim_K$ are

$$KI = KF_1 = \{I, F_1\}, \qquad KR = KF_2 = \{R, F_2\} \qquad \text{and} \qquad KR^2 = KF_3 = \{R^2, F_3\},$$

and the equivalence classes of $\approx_K$ are

$$IK = F_1K = \{I, F_1\}, \qquad RK = F_3K = \{R, F_3\} \qquad \text{and} \qquad R^2K = F_2K = \{R^2, F_2\}.$$

As Lemma 2.12 prescribes, the equivalence classes give partitions of D_3.

2.2.2 Lagrange's theorem

Definition 2.25. Let G be a group and let $x \in G$. The maps

$$\begin{aligned} \ell_x \colon G &\longrightarrow G \\ a &\longmapsto xa \end{aligned} \qquad\qquad \begin{aligned} r_x \colon G &\longrightarrow G \\ a &\longmapsto ax \end{aligned}$$

are called ***left multiplication by*** x and ***right multiplication by*** x, respectively.

Note that these maps are *not* group homomorphisms in general.

Lemma 2.26. *Let G be a group. The maps ℓ_x and r_x defined above are bijections.*

Proof. The inverse of ℓ_x is $\ell_{x^{-1}}$ and the inverse of r_x is $r_{x^{-1}}$. □

Proposition 2.27. *Let G be a group and H a subgroup of G. Then, for any $x \in G$, we have*

$$\#(Hx) = \#H = \#(xH),$$

where # denotes cardinality, that is, the number of elements.

Proof. Let $x \in G$. Then $\ell_x(H) = \{\ell_x(h) \mid h \in H\} = xH$, and thus the restriction of ℓ_x to H is a bijection $H \to xH$. Thus $\#H = \#(xH)$. Similarly, the restriction of r_x to H is a bijection $H \to Hx$, and thus $\#H = \#(Hx)$. □

Recall that, given an equivalence relation, the quotient is the set of equivalence classes. In the case of the relations $\sim_H$ and $\approx_H$, the equivalence classes are the cosets of H, by the last item in Remarks 2.23. More explicitly, the quotients are

$$G/{\sim_H} = \{Hx \mid x \in G\} \qquad \text{and} \qquad G/{\approx_H} = \{xH \mid x \in G\}.$$

Theorem 2.28 (Lagrange's theorem)**.** *Let G be a finite group and $H \subseteq G$ a subgroup. Then*

$$\#(G/{\sim_H}) = \frac{\#G}{\#H} = \#(G/{\approx_H}).$$

Proof. We only show the first equality, which may be rewritten as $\#(G/{\sim_H}) \cdot (\#H) = \#G$. This equality holds because the cosets Hx form a partition of G by Lemma 2.12, the amount of such cosets is $\#(G/{\sim_H})$, and each of them has cardinality equal to $\#H$, by Proposition 2.27. The second equality in the statement can be proven similarly. □

Corollary 2.29. *Let G be a finite group, and denote by e its identity.*

(i) Let H be a subgroup of G. Then $\#H$ divides $\#G$.

(ii) Let $x \in G$. Then the order $\pi(x)$ divides $\#G$.

(iii) If $\#G$ is a prime number, then G is cyclic, and for all $x \neq e$, we have $\langle x \rangle = G$.

(iv) For any $x \in G$, we have $x^{\#G} = e$.

(v) For all $x \in G$, we have $x^{\#G-1} = x^{-1}$.

Proof. Part (i) follows immediately from Lagrange's theorem. As for (ii), by Proposition 1.65, we know that $\pi(x) = \#\langle x \rangle$, and by part (i) we know that $\#\langle x \rangle$ divides $\#G$. Part (iii) is an immediate consequence of (ii). By (ii), we get (iv) because there exists $k \in \mathbb{Z}$ such that $\#G = k\pi(x)$, and then

$$x^{\#G} = x^{k\pi(x)} = (x^{\pi(x)})^k = 1^k = 1.$$

Lastly, (v) follows by multiplying both sides in (iv) by x^{-1}. □

Examples 2.30. Let $p \in \mathbb{Z}$ be a prime number.

- The group $(\mathbb{Z}_p, +)$ has p elements, and we have $\mathbb{Z}_p = \langle \overline{1} \rangle = \langle \overline{n} \rangle$, for any $\overline{n} \neq \overline{0}$.
- The group $(\mathbb{Z}_p \setminus \{\overline{0}\}, \cdot)$ has $p-1$ elements, and therefore $\overline{n}^{p-1} = \overline{1}$ for all $\overline{n} \in \mathbb{Z}_p \setminus \{\overline{0}$.

Remark 2.31. Lagrange's theorem tells us that the cardinality of any subgroup of a finite group G is a divisor of $\#G$. *Warning:* This does not guarantee that for any divisor d of $\#G$ there is a subgroup of G with cardinality equal to d.

2.2.3 Normal subgroups

In the next section, we finally define the quotient of a group by a subgroup, and it turns out that not all subgroups are suitable for this construction. The ones that are, are called *normal subgroups*.

Definition 2.32. Let G be a group. A subgroup H of G is said to be ***normal in*** G if $\sim_H = \approx_H$, which means that $Hx = xH$ for all $x \in G$, or in English words if the left and right coset of H with respect to x are equal, for all x. We write $H \lhd G$ to denote that H is normal in G.

Normal subgroups were first introduced by Evariste Galois in 1831. He used them to decide whether a polynomial is solvable in radicals.

Remarks 2.33. 1. The sets $\{e_G\}$ and G are normal subgroups of G.

2. If G is abelian, then its every subgroup is normal.

Proposition 2.34 (Normality Criterion)**.** *Let G be a group and let $H \subseteq G$ a subgroup. The following are equivalent:*

(1) $H \lhd G$,

(2) $a^{-1}Ha = H$ for all $a \in G$,

(3) $a^{-1}Ha \subseteq H$ for all $a \in G$. (More explicitly, this one means: for all $a \in G$ and for all $h \in H$, we have $a^{-1}ha \in H$.)

Proof. ($1 \Rightarrow 2$) Assume $aH = Ha$ for all $a \in G$. Multiplying with a^{-1} from the left gives $H = a^{-1}Ha$ for all $a \in G$.

($2 \Rightarrow 3$) This implication is immediate.

($3 \Rightarrow 1$) Assume $a^{-1}Ha \subseteq H$ for all $a \in G$. Multiplying by a on the left gives $Ha \subseteq aH$. If $a \in G$, then $a^{-1} \in G$. Hence also $aHa^{-1} \subseteq H$ for all $a \in G$ by the hypothesis. This gives $aH \subseteq Ha$ and hence $Ha = aH$. □

Example 2.35. Consider $D_3 = \{I, R, R^2, F_1, F_2, F_3\}$ and the subgroups $H := \{I, R, R^2\}$ and $K := \{I, F_1\}$. We computed in Example 2.21 the cosets of H and K, and we deduce that H is a normal subgroup, but K is not. This may be visualized by writing the Cayley table of D_3 so that the elements of the same coset of a fixed subgroup occur next to each other. Consider first H and observe that the Cayley table breaks into blocks:

$\circ$	I	R	R^2	F_1	F_2	F_3
I	I	R	R^2	F_1	F_2	F_3
R	R	R^2	I	F_3	F_1	F_2
R^2	R^2	I	R	F_2	F_3	F_1
F_1	F_1	F_2	F_3	I	R	R^2
F_2	F_2	F_3	F_1	R^2	I	R
F_3	F_3	F_1	F_2	R	R^2	$I.$

(The convention here is to write for instance in the row corresponding to F_1 and the column corresponding to R the element $F_1 \circ R$, obtained by performing first R and then F_1.) If instead we take $K := \{I, F_1\}$, then the Cayley table does not break into blocks:

$\circ$	I	F_1	R	F_3	R^2	F_2
I	I	F_1	R	F_3	R^2	F_2
F_1	F_1	I	F_2	R^2	F_3	R
R	R	F_3	R^2	F_2	I	F_1
F_3	F_3	R	F_1	I	F_2	R^2
R^2	R^2	F_2	I	F_1	R	F_3
F_2	F_2	R^2	F_3	R	F_1	$I.$

Recall that the kernel of a homomorphism $f\colon G \to H$ is $\ker(f) := \{g \in G \mid f(g) = e_H\}$. We showed in Proposition 1.79 that $\ker(f)$ is a subgroup of G.

Proposition 2.36. *For any group homomorphism $f\colon G \to H$, the kernel $\ker(f)$ is a normal subgroup of G.*

Proof. We need to prove that $a^{-1}\ker(f)a \subseteq \ker(f)$ for all $a \in G$. Let $x \in a^{-1}\ker(f)a$. Then $x = a^{-1}ya$ for some $y \in \ker(f)$. We compute

$$f(x) = f(a^{-1}ya) = f(a^{-1})f(y)f(a) = f(a)^{-1}e_H f(a) = e_H.$$

Hence $x \in \ker(f)$, which proves the claim. □

Remark 2.37. The image of a homomorphism is not necessarily normal. For instance the image of the homomorphism

$$\mathbb{Z}_2 \longrightarrow S_3, \qquad \overline{0} \longmapsto \mathrm{id}_{S_3}, \quad \overline{1} \longmapsto (12)$$

is the subgroup $\langle (12) \rangle \subset S_3$. This is not a normal subgroup, for instance because $(13)(12)(13) = (23) \notin \langle (12) \rangle$.

Definition 2.38. A group G is called a ***simple group*** if its normal subgroups are $\{e_G\}$ and G.

Example 2.39. If $p \in \mathbb{Z}$ is prime, the cyclic group $\mathbb{Z}_p$ is a simple group.

Proposition 2.40. *Let G be a finite group and let H be a subgroup with $\frac{\#G}{\#H} = 2$. Then $H \lhd G$.*

Proof. You may check as an exercise that:

- for all $h \in H$, we have $Hh = H = hH$,
- for all $g \in G \setminus H$, we have $Hg = G \setminus H = gH$.

In particular, the left and right cosets coincide. □

2.2.4 Group structure on a quotient

The general slogan is: In order to define a group structure on the quotient $G/{\sim_H}$ (or $G/{\approx_H}$) in a natural way, H has to be a normal subgroup. Recall that if H is a normal subgroup, we have by definition ${\sim_H} = {\approx_H}$, which means that $Hx = xH$ for all $x \in G$.

Definition 2.41. Let G be a group and let $H \lhd G$. We define

$$G/H := G/\sim_H = \{Hx \mid x \in G\}$$

the ***quotient of*** G ***by*** H, or ***modulo*** H, and on G/H we define the binary operation

$$\begin{aligned} \cdot : G/H \times G/H &\longrightarrow G/H \\ (Hx, Hy) &\longmapsto Hxy. \end{aligned}$$

We read G/H as "G over H" or "G modulo H" (but not "G divided by H").

Lemma 2.42. *The operation on G/H is well-defined, that is, if $Hx = Hx'$ and $Hy = Hy'$, then $Hxy = Hx'y'$.*

Proof. The equality $Hx = Hx'$ implies that $x = ax'$ for some $a \in H$, and similarly we have $y = by'$ for some $b \in H$. We then compute

$$Hxy = Hax'by' = Hx'by' = x'Hby' = x'Hy' = Hx'y',$$

where in the second and fourth equalities we use the fact that $hH = Hh = H$ for any $h \in H$, which holds because H is a subgroup, and in the third and fourth equalities we use the fact that H is assumed to be normal, and hence $Hx' = x'H$. □

Theorem 2.43. *Let G be a group and $H \lhd G$. The operation $\cdot$ defined above induces a group structure on G/H. Furthermore, there is a natural surjective homomorphism*

$$\begin{aligned} p_H \colon G &\longrightarrow G/H, \\ x &\longmapsto Hx. \end{aligned}$$

Moreover, the kernel of p_H is H.

Proof. We have already checked that $\cdot$ is a well-defined binary operation. We will check the properties in the definition of a group:

- Associativity: Let $x, y, z \in G$. Then

  $$(Hx)\cdot(Hy \cdot Hz) = (Hx)\cdot(Hyz) = H(x(yz)) = H((xy)z) = (Hxy)\cdot(Hz) = (Hx \cdot Hy)\cdot(Hz),$$

  where in the middle equality we used the associativity in G.

- Identity element: The element He is the identity element of G/H, since

  $$Hx \cdot He = Hxe = Hx = Hex = He \cdot Hx$$

  for all $x \in G$. We used the fact that e is the identity in G.

- Inverses: For $x \in G$, the inverse of Hx in G/H is Hx^{-1}, since

  $$Hx \cdot Hx^{-1} = Hxx^{-1} = He = Hx^{-1}x = Hx^{-1} \cdot Hx.$$

Hence G/H is a group. It is easy to see that p_H is a surjective homomorphism. Lastly, $\ker(p_H)$ consists of all $x \in G$ such that $Hx = H$. Hence $\ker(p_H) = H$. □

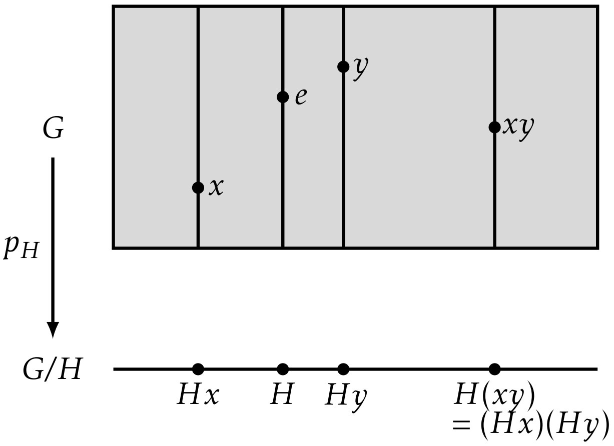


Figure 2.1: Cosets of H collapsed by p_H.

Examples 2.44. 1. Let $G = \{I, R, R^2, F_1, F_2, F_3\}$ and $H = \{I, R, R^2\}$. We already know that $H \lhd G$, and $\frac{\#G}{\#H} = 2$. As we observed in the proof of Proposition 2.40, the cosets are H and $G \setminus H = \{F_1, F_2, F_3\}$. The Cayley table of G/H is

	H	$G \setminus H$
H	H	$G \setminus H$
$G \setminus H$	$G \setminus H$	H,

from which it is clear that $G/H \cong \mathbb{Z}_2$. And this is true in general whenever $\frac{\#G}{\#H} = 2$.

2. Consider the modulus map from the $(\mathbb{C} \setminus \{0\}, \cdot)$ to the group $(\mathbb{R}_{>0}, \cdot)$. This is a homomorphism, since for complex numbers z_1 and z_2 we have $|z_1 z_2| = |z_1||z_2|$. Since $\{1\}$ is a subgroup of $\mathbb{R}_{>0}$, its preimage U, which is the set of complex numbers of modulus 1, is a subgroup of $\mathbb{C}^*$. The modulus of a complex number is its distance from the origin. Hence the cosets of U are circles around the origin. Each circle is mapped by this homomorphism to its intersection with the positive real axis.

2.2.5 Quotients of $\mathbb{Z}$

Let $G = \mathbb{Z}$ under addition and $H = n\mathbb{Z}$. Then H is normal, because G is abelian. We have

$$\mathbb{Z}/n\mathbb{Z} = \{a + n\mathbb{Z} \mid a \in \mathbb{Z}\}.$$

Two cosets $a + n\mathbb{Z}$ and $a' + n\mathbb{Z}$ are the same if $n|(a - a')$, or equivalently if $a \equiv a' \mod n$. Hence

$$\mathbb{Z}/n\mathbb{Z} = \{a + n\mathbb{Z} \mid a = 0, 1, \ldots, n-1\}.$$

As a matter of fact, the operation on $\mathbb{Z}/n\mathbb{Z}$ involves integer addition and then reduction mod n:

$$(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z},$$

and the identity is $n\mathbb{Z} = 0 + n\mathbb{Z}$. The quotient group $\mathbb{Z}/n\mathbb{Z}$ is isomorphic to $\mathbb{Z}_n$, an isomorphism being

$$\begin{aligned} \mathbb{Z}/n\mathbb{Z} &\longrightarrow \mathbb{Z}_n \\ a + n\mathbb{Z} &\longmapsto [a]_n. \end{aligned}$$

Some authors take a simple approach to $\mathbb{Z}_n$, saying that it is the set of symbols of the form $[0]_n, [1]_n, \dots, [n-1]_n$, equipped with addition modulo n. Elsewhere in the literature, the group $\mathbb{Z}_n$ is defined directly as $\mathbb{Z}/n\mathbb{Z}$.

The terminology from $\mathbb{Z}/n\mathbb{Z}$ is often carried over to factor groups. A factor group G/H is called the *factor group of G modulo H*. Elements in the same coset of H are called *congruent modulo H*.

2.3 Isomorphism theorems for groups

Notation 2.45. From now on, when G is a group and $H \lhd G$, for any $x \in G$ we denote the class of x in G/H by $\overline{x}$. That is, we write

$$\overline{x} := xH = Hx.$$

If there is ambiguity about the quotient, we will still write cosets explicitly.

2.3.1 First isomorphism theorem for groups

Lemma 2.46. *Let G be a group and $H \lhd G$. The* **projection** $p\colon G \to G/H$ *defined by* $p(x) := \overline{x}$ *is a group homomorphism.*

Proof. For any $x, y \in G$, we have

$$p(xy) = \overline{xy} = \overline{x} \cdot \overline{y} = p(x) \cdot p(y).$$

□

Recall that the kernel of a homomorphism is a normal subgroup, by Proposition 2.36, hence we can form the quotient by it.

Lemma 2.47. *Let $f\colon G \to H$ be a homomorphism. Consider the relation $\sim_f$ introduced in Definition 2.14 and the relation $\sim_{\ker(f)}$ in Definition 2.22. These relations are the same, or in other words the quotients $G/{\sim_f}$ and $G/\ker(f)$ are the same.*

Proof. We need to show that, for any $x, y \in G$, we have $x \sim_f y$ if and only if $x \sim_{\ker(f)} y$. By definition, $x \sim_f y$ means that $f(x) = f(y)$, which is equivalent to

$$f(yx^{-1}) = f(y)f(x^{-1}) = f(y)f(x)^{-1} = e_H.$$

This means that $yx^{-1} \in \ker(f)$, which is equivalent to $x \sim_{\ker(f)} y$, by the second item of Remark 2.23. □

Notation 2.48. From now on, we also omit the $\cdot$ in the quotient, writing the operation in the quotient simply as juxtaposition.

Theorem 2.49 (First Isomorphism Theorem). *Let $f\colon G \to H$ be a group homomorphism. There is an isomorphism*

$$G/\ker(f) \cong \operatorname{im}(f).$$

More precisely, consider the function $\varphi\colon G/\ker(f) \to H$ defined by $\varphi(\overline{x}) := f(x)$. The function φ is an injective group homomorphism that makes the diagram

$$\begin{tikzcd}
G \arrow[r, "f"] \arrow[d, "p"'] & H \\
G/\ker(f) \arrow[ur, dashed, "\varphi"'] &
\end{tikzcd}$$

commute (i.e., $\varphi \circ p = f$). In particular, φ induces the isomorphism $G/\ker(f) \cong \operatorname{im}(f)$.

Proof. By Lemma 2.47, we know that $G/\ker(f)$ is equal to the quotient $G/\sim_f$ considered in Theorem 2.16, which automatically guarantees that:

- the function φ is well-defined and injective,
- the diagram in the statement commutes, that is, $\varphi \circ p = f$,
- the function φ induces a bijection $G/\ker(f) \to \operatorname{im}(f)$.

So we only need to show that φ is a group homomorphism: for any $x, y \in G$, we have

$$\varphi(\overline{x}\,\overline{y}) = \varphi(\overline{xy}) = f(xy) = f(x)f(y) = \varphi(\overline{x})\varphi(\overline{y}),$$

where the third equality holds because f is a homomorphism. □

Examples 2.50. 1. Let G be any group. The identity homomophism $\mathrm{id}_G \colon G \to G$ is such that $\ker(f) = \{e_G\}$. The first isomorphim theorem then states that

$$G/\{e_G\} \cong G.$$

2. On the opposite extreme, for any two groups G and H consider the trivial homomorphism $f\colon G \to H$ given by $f(x) := e_H$ for all $x \in G$, where e_H is the identity of H. In this case $\ker(f) = G$, and by the first isomorphism theorem we get

$$G/G \cong \{e_H\}.$$

3. Specialized to $\mathbb{Z}$, the two examples above give

$$\mathbb{Z}/\{0\} \cong \mathbb{Z} \qquad \text{and} \qquad \mathbb{Z}/\mathbb{Z} \cong \{0\}.$$

4. Fix $n \in \mathbb{Z}_{>0}$. The homomorphism

$$\begin{aligned} f\colon \mathbb{Z} &\longrightarrow \mathbb{Z}_n \\ x &\longmapsto \overline{x} \end{aligned}$$

is surjective, and $\ker(f) = n\mathbb{Z}$. Hence $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.

5. Consider the homomorphism

$$\begin{aligned} f\colon (\mathbb{R},+) &\longrightarrow \big(\mathbb{C}\setminus\{0\},\cdot\big) \\ \alpha &\longmapsto \cos(2\pi\alpha) + i\sin(2\pi\alpha). \end{aligned}$$

It is a homomorphism: $f(\alpha+\beta) = f(\alpha)f(\beta)$. The kernel is $\ker(f) = \mathbb{Z}$, and the image is $\operatorname{im}(f) = \{z \in \mathbb{C} \mid |z| = 1\}$. Hence

$$\mathbb{R}/\mathbb{Z} \cong \{z \in \mathbb{C} \mid |z| = 1\}.$$

6. Consider the homomorphism

$$\begin{aligned} \det\colon \mathrm{GL}_n(\mathbb{C}) &\longrightarrow \big(\mathbb{C}\setminus\{0\},\cdot\big) \\ A &\longmapsto \det A. \end{aligned}$$

Then $\ker(\det) = \mathrm{SL}_n(\mathbb{C}) := \{A \in \mathrm{GL}_n(\mathbb{C}) \mid \det(A) = 1\}$, and the image is the whole $\mathbb{C}\setminus\{0\}$. Therefore

$$\mathrm{GL}_n(\mathbb{C})/\mathrm{SL}_n(\mathbb{C}) \cong \mathbb{C}\setminus\{0\},$$

from which in particular we deduce that this quotient is abelian and infinite.

2.3.2 Second isomorphism theorem for groups

Recall that for two subsets H and K of a group G we denote

$$HK := \{ab \mid a \in H, b \in K\}.$$

Lemma 2.51. *Let G be a group, and let H and N be subgroups of G, with $N \lhd G$. Then*

(1) $HN = NH$,

(2) $\langle H \cup N \rangle = HN$.

Proof. (1) We show that $HN \subseteq NH$. Let $a \in H$ and $b \in N$. Then $ab \in aN = Na \subseteq NH$. The converse inclusion is analogous.

(2) We need to show that HN is a subgroup of G and that it is contained in any subgroup T of G that contains $H \cup N$. We check that HN is a subgroup by using the subgroup criterion:

- We have $e_G \in H$ and $e_G \in N$, so that $e_G = e_G e_G \in HN$, and hence $HN \neq \emptyset$.
- For all $ab, cd \in HN$, we have

$$ab(cd)^{-1} = abd^{-1}c^{-1} \in HNNH = HNH = HHN = HN.$$

Let now T be a subgroup of G such that $H \cup N \subseteq T$. This means that $H \subseteq T$ and $N \subseteq T$, but then, since T is a subgroup, we have $HN \subseteq T$. □

Remark 2.52. Let G be a group, and let H and N be subgroups of G, with $N \lhd G$. We saw in the previous lemma that

$$HN = \{ab \mid a \in H, b \in N\}$$

is a subgroup of G. Moreover we observe that N is a normal subgroup in HN (by the normality in G), and $H \cap N$ is a normal subgroup of H.

Theorem 2.53 (Second Isomorphism Theorem)**.** *Let G be a group, and let H and N be subgroups of G, with $N \lhd G$. Then*

$$(HN)/N \cong H/(H \cap N).$$

Proof. Consider the homomorphism $\Phi\colon H \to HN/N$ obtained as the composition of

$$\begin{aligned} H &\longrightarrow HN \\ h &\longmapsto h \end{aligned} \qquad\qquad \begin{aligned} HN &\longrightarrow HN/N \\ h &\longmapsto \overline{h}, \end{aligned}$$

that is, $\Phi(h) := \overline{h}$. Then Φ is a group homomorphism. Let us show that it is surjective: Let $ab \in HN$, so that $\overline{ab} = \overline{a}$ in HN/N (since $b \in N$). But then $\overline{ab} = \Phi(a)$, and therefore Φ is surjective. Now we observe that

$$\ker(\Phi) = \{h \in H \mid \overline{h} = \overline{e}\} = \{h \in H \mid h \in N\} = H \cap N.$$

The statement then follows from the first isomorhism theorem. □

In additive notation, the isomorphism is written as

$$(H+N)/N \cong H/(H \cap N).$$

Examples 2.54.

- Consider $G = \mathbb{Z}$, and let $m, n \in \mathbb{Z}$. If we define $H := \langle m \rangle = m\mathbb{Z}$ and $N := \langle n \rangle = n\mathbb{Z}$, we get

$$\begin{aligned} H + N &= \langle m, n \rangle = \langle \gcd(m,n) \rangle = \gcd(m,n)\mathbb{Z} \\ H \cap N &= \langle \operatorname{lcm}(m,n) \rangle = \operatorname{lcm}(m,n)\mathbb{Z}. \end{aligned}$$

  Then the second isomorphism theorem states that

$$\frac{\gcd(m,n)\mathbb{Z}}{n\mathbb{Z}} \cong \frac{m\mathbb{Z}}{\operatorname{lcm}(m,n)\mathbb{Z}}.$$

- Consider $G = \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ with coordinate-wise addition. Define $H := \mathbb{Z} \times \mathbb{Z} \times \{0\}$, and $N := \{0\} \times \mathbb{Z} \times \mathbb{Z}$. Then

$$\begin{aligned} H + N &= \mathbb{Z} \times \mathbb{Z} \times \mathbb{Z} \\ H \cap N &= \{0\} \times \mathbb{Z} \times \{0\} \end{aligned}$$

  We have $(H + N)/N \cong \mathbb{Z}$ and also $H/(H \cap N) \cong \mathbb{Z}$.

2.3.3 Third isomorphism theorem for groups

Lemma 2.55. *Let G be a group, and let H and K be subgroups, with $K \lhd G$ and $K \subseteq H \subseteq G$. Then:*

(1) $K \lhd H$,

(2) the set $H/K = \{hK \mid h \in H\}$ is a subgroup of $G/K = \{gK \mid g \in G\}$,

(3) if in addition $H \lhd G$, then $H/K \lhd G/K$.

Proof. (1) This is clear. (2) Exercise. (3) Let $hK \in H/K$ and $gK \in G/K$. Then

$$gK\, hK\, (gK)^{-1} = (ghg^{-1})K \in H/K,$$

where the membership holds because $ghg^{-1} \in H$, since $H \lhd G$. □

Remark 2.56. In the hypotheses of the above lemma, including (3), the subgroup H/K is a normal subgroup of G/K, hence we may form the quotient $(G/K)\big/(H/K)$. As it happens with the quotient of any group, the elements the quotient are cosets of the subgroup we quotient by, so that the elements of *this* big quotient are cosets of H/K. A bit more explicitly,

$$\begin{aligned} (G/K)\big/(H/K) &= \Big\{(H/K)gK \mid gK \in G/K\Big\} \\ &= \Big\{\{hK \mid h \in H\}gK \mid gK \in G/K\Big\} \\ &= \Big\{\{hgK \mid h \in H\} \mid g \in G\Big\}. \end{aligned}$$

Theorem 2.57 (Third Isomorphism Theorem)**.** *Let G be a group and let H and K be normal subgroups of G, with $K \subseteq H$. Then*

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

Proof. Consider the map

$$\begin{aligned} f\colon G/K &\to G/H \\ aK &\mapsto aH. \end{aligned}$$

We will show that this map is well-defined, a homomorphism, surjective and that its kernel is equal to H/K.

- Well-definedness: Let $aK = bK$ for $a, b \in G$. Then there exists $k \in K$ such that $a = bk$. Hence

  $$f(aK) = aH = (bk)H = (bH)(kH) = (bH)(eH) = (be)H = bH = f(bK),$$

  where the third equality follows from $k \in H$.

- Homomorphism: Let $a, b \in G$. Then

  $$f((aK)(bK)) = f((ab)K) = (ab)H = aHbH = f(aK)f(bK).$$

- Surjectivity: Let $x \in G/H$. Then $x = aH$ for $a \in G$. The map f is surjective, because $f(aK) = aH = x$.

- Kernel: The kernel of f is

  $$\ker(f) = \{aK \mid a \in G, aH = H\} = \{aK \mid a \in H\} = H/K.$$

Now the statement follows from the first isomorphism theorem. □

Example 2.58. Consider $G = \mathbb{Z}$. Let $H := 8\mathbb{Z}$ and $K := 16\mathbb{Z}$. By the third isomorphism theorem, we get

$$(\mathbb{Z}/16\mathbb{Z})\Big/(8\mathbb{Z}/16\mathbb{Z}) \cong \mathbb{Z}/8\mathbb{Z}.$$

2.3.4 The correspondence theorem

Theorem 2.59 (Correspondence Theorem)**.** *Let G be a group, and let $K \lhd G$. The functions*

$$\begin{aligned} \Phi\colon \{\textit{subgroups of } G \textit{ containing } K\} &\longrightarrow \{\textit{subgroups of } G/K\} \\ H &\longmapsto H/K \end{aligned}$$

and

$$\begin{aligned} \Psi\colon \{\textit{normal subgroups of } G \textit{ containing } K\} &\longrightarrow \{\textit{normal subgroups of } G/K\} \\ H &\longmapsto H/K \end{aligned}$$

are bijections.

Proof. Both maps are well-defined by Lemma 2.55. We show that Φ is a bijection:

- (Injectivity) Let H_1 and H_2 be subgroups with $K \subseteq H_i \subseteq G$, and such that $H_1/K = H_2/K$. Then we show that $H_1 = H_2$.

  ($\subseteq$) Let $x \in H_1$. Then $xK \in H_1/K = H_2/K$, which means that there exists $y \in H_2$ such that $xK = yK$. But then $xy^{-1} \in K$, so that $x \in Ky \subseteq H_2$, and hence $x \in H_2$.

  ($\supseteq$) Same thing, changing the roles of H_1 and H_2.

- (Surjectivity) Let $A \subseteq G/K$ be a subgroup. Define $H := \{g \in G \mid gK \in A\}$. We have to show that $K \subseteq H$, that H is a subgroup of G, and that $\Phi(H) = A$.
  - Let $k \in K$. Then $kK = eK \in A$, since A is a subgroup, and therefore $k \in H$. Hence $K \subseteq H$.
  - Since $1 \in K \subseteq H$, we know that $H \neq \emptyset$. Moreover, given $x, y \in H$, by definition we know that $xK \in A$ and $yK \in A$, and since A is a subgroup we know that $xK,(yK)^{-1} = (xy^{-1})K \in A$, which means that $xy^{-1} \in H$. Hence H is a subgroup of G, by the subgroup criterion.
  - We have $\Phi(H) = H/K = \{xK \mid x \in H\} = \{xK \mid xK \in A\} = A$.

Next we see that Ψ is a bijection. It is injective, because we simply restricted domain and codomain of Φ, which is a bijective function. To prove the surjectivity of Ψ, let $A \lhd G/K$. Construct $H := \{g \in G \mid gK \in A\}$. Similarly to the part about Φ, we may see that $\Psi(H) = A$. The only thing left to check is that H is normal. Indeed, for all $x \in H$ and $g \in G$, we have

$$(gxg^{-1})K = gK\,xK\,(gK)^{-1} \in A,$$

where the membership holds because $xK \in A$ and A is assumed to be normal, and this means that $gxg^{-1} \in H$. □

Example 2.60. Consider $G = \mathbb{Z}$ and the subgroup $K = \langle 12 \rangle = 12\mathbb{Z}$. Then $G/K \cong \mathbb{Z}_{12}$. We wish to describe the subgroups of $\mathbb{Z}_{12}$. By the correspondence theorem, there is a bijection

$$\{\text{subgroups of } \mathbb{Z}_{12}\} \longleftrightarrow \left\{H \mid 12\mathbb{Z} \subseteq H \overset{\text{subg.}}{\subseteq} \mathbb{Z}\right\}.$$

More explicitly, the bijection tells us that

$$\{\text{subgroups of } \mathbb{Z}_{12}\} = \left\{H/12\mathbb{Z} \mid 12\mathbb{Z} \subseteq H \overset{\text{subg.}}{\subseteq} \mathbb{Z}\right\}.$$

The subgroups of $\mathbb{Z}$ are of the form $\langle n \rangle = n\mathbb{Z}$, for $n \in \mathbb{Z}$. And we have $12\mathbb{Z} \subseteq n\mathbb{Z}$ if and only if n divides 12. The divisors of 12 are: 1, 2, 3, 4, 6 and 12. The subgroups of $\mathbb{Z}_{12}$ are therefore

$$\begin{array}{lll} \mathbb{Z}_{12} = \langle \overline{1} \rangle \cong \mathbb{Z}/12\mathbb{Z} & \langle \overline{3} \rangle \cong 3\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}_4 & \langle \overline{6} \rangle \cong 6\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}_2 \\ \quad \langle \overline{2} \rangle \cong 2\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}_6 & \langle \overline{4} \rangle \cong 4\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}_3 & \{\overline{0}\} = \langle \overline{12} \rangle \cong 12\mathbb{Z}/12\mathbb{Z}. \end{array}$$

The following diagrams show inclusions of subgroups, respectively in $\mathbb{Z}$ and in $\mathbb{Z}_{12} = \mathbb{Z}/12\mathbb{Z}$:

$$\begin{tikzcd}
 & \mathbb{Z} & \\
2\mathbb{Z} \arrow[ur, hook] & & 3\mathbb{Z} \arrow[ul, hook] \\
4\mathbb{Z} \arrow[u, hook] & & 6\mathbb{Z} \arrow[u, hook] \arrow[ull, hook] \\
 & 12\mathbb{Z} \arrow[ul, hook] \arrow[ur, hook] &
\end{tikzcd}
\qquad
\begin{tikzcd}
 & \mathbb{Z}/12\mathbb{Z} & \\
2\mathbb{Z}/12\mathbb{Z} \arrow[ur, hook] & & 3\mathbb{Z}/12\mathbb{Z} \arrow[ul, hook] \\
4\mathbb{Z}/12\mathbb{Z} \arrow[u, hook] & & 6\mathbb{Z}/12\mathbb{Z} \arrow[u, hook] \arrow[ull, hook] \\
 & 12\mathbb{Z}/12\mathbb{Z} \arrow[ul, hook] \arrow[ur, hook] &
\end{tikzcd}$$

Chapter 3

Some important groups and results

We start the chapter by studying in greater detail the symmetric group S_n. We then introduce a construction that given a family of groups builds a larger group, their *direct product*, in Section 3.2. After that we describe:

- all groups of size ≤ 8, in Section 3.3,
- all finitely generated abelian groups, in Section 3.4.

3.1 The symmetric group

With the First Isomorphism Theorem in our toolbox, we can now prove a very important result: Cayley's theorem.

Recall that in Definition 1.34, given a set X, we introduced the ***symmetric group of*** X as

$$S_X := \{\text{invertible functions } X \to X\}.$$

In the special case of $X = \{1, \dots, n\}$, we denote this by

$$S_n := S_{\{1,\dots,n\}}$$

and we call it the n***-th symmetric group***.

3.1.1 Cayley's theorem

Lemma 3.1. *If a set X has n elements, then the group S_X is isomorphic to S_n.*

Proof. Exercise. □

For the proof of the following theorem to make sense, the operation on S_X is

$$\begin{aligned} S^X \times S^X &\longrightarrow S^X \\ (g,h) &\longmapsto g \circ h. \end{aligned}$$

If instead one defines the operation by swapping the two functions, tha is, $h \circ g$, the proof below can be modified by taking $\ell_{x^{-1}}$.

Theorem 3.2 (Cayley's theorem)**.** *Every finite group is isomorphic to a subgroup of a symmetric group. More precisely, let G be a group with $\#G = n$. Then G is isomorphic to a subgroup of S_n.*

Proof. For $x \in G$, consider the function $\ell_x \colon G \to G$ defined by $\ell_x(a) := xa$. This function is a bijection by Lemma2.26. We claim that the map

$$\begin{aligned} f \colon G &\longrightarrow S_G \\ x &\longmapsto \ell_x \end{aligned}$$

is an injective homomorphism. The map f is injective, since $\ell_x(e) = xe = x$ and hence $\ell_x \neq \ell_y$ for $x \neq y$. The map f is a homomorphism, because

$$\ell_{xy}(a) = (xy)a = x(ya) = \ell_x(ya) = \ell_x(\ell_y(a)) = (\ell_x \circ \ell_y)(a)$$

for all $a \in G$, and hence

$$f(x) \circ f(y) = \ell_x \circ \ell_y = \ell_{xy} = f(xy).$$

But then f induces an isomorphism between G and $\operatorname{im}(f)$, and $\operatorname{im}(f)$ is a subgroup of S_G. Lastly, since S_G is isomorphic to S_n by Lemma 3.1, it follows that G is isomorphic to a subgroup of S_n. □

3.1.2 Decomposing permutations into cycles

Recall that an element $\sigma \in S_n$ may be written as

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

and that we simply juxtapose elements, namely we write $\sigma\tau$ instead of $\sigma \circ \tau$. For instance

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

Definition 3.3. Let $\sigma \in S_n$. If there are r elements $a_1, \dots, a_r \in \{1, \dots, n$
such that

$$\sigma(a_1) = a_2, \quad \sigma(a_2) = a_3, \quad \sigma(a_3) = a_4, \quad \dots, \quad \sigma(a_{r-1}) = a_r, \quad \sigma(a_r) = a_1$$

with $a_i \neq a_j$ for $i \neq j$, and for all elements $b \in \{1, \dots, n\} \setminus \{a_1, \dots, a_n\}$ we have $\sigma(b) = b$, then we call σ an r-***cycle***, or a ***cycle of order*** r. When r is not important, we simply call σ a ***cycle***. We denote this r-cycle σ by $(a_1, a_2, \dots, a_r)$.

Example 3.4. The permutation $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ is a 3-cycle, and in cycle notation it may be written as $(1,2,3)$, or alternatively as $(2,3,1)$ or $(3,1,2)$. The permutations

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

are 2-cycles, and in cycle notation we may write them as $(2,3)$ and $(1,3)$, respectively. Their composition, written in cycle notation, is $(2,3)(1,3) = (1,2,3)$.

Remark 3.5. Note that the group-theoretic order of an r-cycle is exactly r.

Definition 3.6. We say that two cycles in S_n are disjoint if any given $i \in \{1, \dots, n\}$ is changed by at most one of the two cycles.

For instance the two 2-cycles in the previous example are not disjoint, because 3 is changed by both.

Proposition 3.7. *Any permutation $\sigma \in S_n$ may be expressed as a product of disjoint cycles. This factorization is unique, ignoring 1-cycles, up to the order of the factors.*

Proof. We construct a partition $\{1,\ldots,n\} = A_1 \cup A_2 \cup \ldots \cup A_r$ as follows: we put two elements a and b of $\{1,\ldots,n\}$ in the same set A_i if $b = \sigma^k(a)$ for some $k \in \mathbb{Z}$. This is indeed a partition of $\{1,\ldots,n\}$, and one way to see this is by showing that the relation on $\{1,\ldots,n\}$ defined by setting $a \sim b$ if $b = \sigma^k(a)$ for some $k \in \mathbb{Z}$ is an equivalence relation:

- (Reflexivity) We have $a \sim a$, since $a = \sigma^0(a)$.
- (Symmetry) Let $a \sim b$. Then there exists $k \in \mathbb{Z}$ such that $b = \sigma^k(a)$. But then $a = \sigma^{-k}(b)$ and $-k \in \mathbb{Z}$, thus $b \sim a$.
- (Transitivity) Let $a \sim b$ and $b \sim c$. Then there exist $k,\ell \in \mathbb{Z}$ such that $b = \sigma^k(a)$ and $c = \sigma^\ell(b)$. Thus $c = \sigma^\ell(b) = \sigma^\ell(\sigma^k(a)) = \sigma^{k+\ell}(a)$, which means that $a \sim c$.

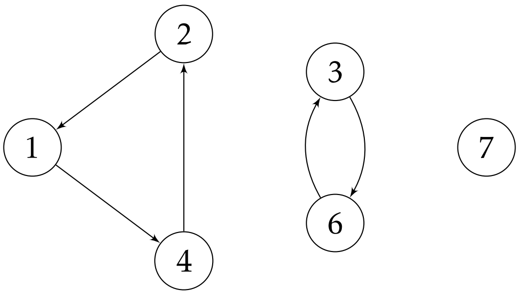
The A_i's are then defined as the equivalence classes of $\sim$. Define now the cycle μ_i as

$$\mu_i(x) := \begin{cases} \sigma(x) & \text{if } x \in A_i, \\ x & \text{otherwise.} \end{cases}$$

Since the equivalence classes $A_1, A_2, \ldots, A_k$ are disjoint, the cycles $\mu_1, \mu_2, \ldots, \mu_k$ are disjoint as well. And $\sigma = \mu_1 \mu_2 \ldots \mu_r$.

Now for the uniqueness: Assume there is another factorization $\sigma = \tau_1 \tau_2 \ldots \tau_s$ into disjointed cycles. Because of the disjointness, for any $a \in \{1,\ldots,n\}$, there is exactly one cycle acting on a, either as the identity or changing a. But then we see that the τ_i's have to act on each element the same way σ does, hence the two factorizations are the same, up to the order. □

Remark 3.8. One may visualize a permutation as a directed graph: the vertices are $1, 2, \ldots, n$, and there is exactly one arrow shooting out and exactly one arrow shooting in every vertex. Two numbers are in the same cycle exactly if the corresponding vertices are in the same connected component of the graph (which is a cycle in the graph-theoretic sense).



Example 3.9. The permutation depicted in the previous remark may be factored as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 1 & 6 & 2 & 5 & 3 & 7 \end{pmatrix} = (1,4,2)(3,6)(7) = (6,3)(7)(1,4,2).$$

Definition 3.10. The 2-cycles are called ***transpositions***.

Corollary 3.11. *The transpositions generate S_n.*

Proof. Since every permutation is a product of cycles by Proposition 3.7, it is enough to show that any cycle is a product of transpositions: $(a_1, a_2, \ldots, a_r) = (a_1, a_r)(a_1, a_{r-1}) \ldots (a_1, a_2)$. □

3.1.3 The alternating group

Definition 3.12. For $\sigma \in S_n$, we let $C(\sigma) := \{(i,j) \mid 1 \le i < j \le n, \sigma(i) > \sigma(j)\}$, and define the ***sign of*** σ to be the number

$$\operatorname{sign}(\sigma) := (-1)^{\#C(\sigma)} = \begin{cases} 1 & \text{if } \#C(\sigma) \text{ is even,} \\ -1 & \text{if } \#C(\sigma) \text{ is odd.} \end{cases}$$

We call σ an ***even permutation*** if $\operatorname{sign}(\sigma) = 1$, or an ***odd permutation*** if $\operatorname{sign}(\sigma) = -1$.

Example 3.13. In any S_n, the transpositions are odd, and the identity is even.

Proposition 3.14. *The function*

$$\begin{aligned} \operatorname{sign}\colon S_n &\longrightarrow \big(\{1,-1\},\cdot\big) \\ \sigma &\longmapsto \operatorname{sign}(\sigma) \end{aligned}$$

is a surjective group homomorphism.

Proof. Exercise. □

Definition 3.15. The kernel of the sign map is called the ***alternating group*** A_n.

By Proposition 2.36, we know that the kernel of a homomorphism is a normal subgroup, hence $A_n \lhd S_n$, and we may form the quotient S_n/A_n. By the first isomorphism theorem, we have

$$S_n/A_n \cong \{1,-1\} \cong \mathbb{Z}_2.$$

Examples 3.16.

- The group $A_3 = \{\mathrm{id}_{\{1,2,3\}}, (1,2,3), (1,3,2)\}$ is abelian and cyclic, and more explicitly we have $A_3 \cong \mathbb{Z}_3$.
- The group

$$\begin{aligned} A_4 = \Big\{&\mathrm{id}_{\{1,2,3,4\}}, \\ &(1,2,3), (1,3,2), (1,2,4), (1,4,2), (1,3,4), (1,4,3), (2,3,4), (2,4,3), \\ &(1,2)(3,4), (1,3)(2,4), (1,4)(2,3)\Big\} \end{aligned}$$

is not abelian, and it does not have subgroups of 6 elements. Exercise.

3.2 Direct product of groups

Definition 3.17. Given two groups $(G,*_G)$ and $(H,*_H)$, we define a group structure on the Cartesian product $G \times H = \{(g,h) \mid g \in G, h \in H\}$, by setting

$$(g,h) * (g',h') := (g *_G g', h *_H h')$$

for all $(g,h),(g',h') \in G \times H$. The group $(G \times H, *)$ is called the ***direct product of*** $(G,*_G)$ ***and*** $(H,*_H)$.

The elements of $G \times H$ behave like vectors in linear algebra: the operation is performed component-wise.

Proposition 3.18. *The definition above does indeed yield a group.*

Proof. You may check as an exercise that: (1) associativity holds, (2) the identity is (e_G, e_H), (3) the inverse of (g,h) is (g^{-1}, h^{-1}). □

Examples 3.19. 1. From two copies of $\mathbb{Z}$ under addition we may form the direct product $\mathbb{Z}\times\mathbb{Z}$. In this group we have for instance $(1,1)+(4,6)=(5,7)$ and $(2,3)+(-1,0)=(1,3)$. Observe that $\mathbb{Z}\times\mathbb{Z}$ is not cyclic: for any $(a,b)\in\mathbb{Z}\times\mathbb{Z}$, the subgroup generated by (a,b) is $\langle(a,b)\rangle=\{(na,nb)\mid n\in\mathbb{Z}\}$ and in particular if the signs of a and b are the same, then also the components of every element in $\langle(a,b)\rangle$ have the same sign (so that for instance $(1,-1)\notin\langle(a,b)\rangle$ in this case), and if the signs of a and b are different, then also the components of every element in $\langle(a,b)\rangle$ have different sign (so that for instance $(1,1)\notin\langle(a,b)\rangle$ in this case).

2. The group $\mathbb{Z}_2\times\mathbb{Z}_2$ has four elements, and its Cayley table is

	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$
$(\overline{0},\overline{0})$	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$
$(\overline{0},\overline{1})$	$(\overline{0},\overline{1})$	$(\overline{0},\overline{0})$	$(\overline{1},\overline{1})$	$(\overline{1},\overline{0})$
$(\overline{1},\overline{0})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$
$(\overline{1},\overline{1})$	$(\overline{1},\overline{1})$	$(\overline{1},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{0},\overline{0})$.

We know another group with four elements, $\mathbb{Z}_4$. These two groups are *not* isomorphic, since $\mathbb{Z}_4$ contains an element of order 4, but $\mathbb{Z}_2\times\mathbb{Z}_2$ does not contain any such element. (And by Corollary 1.82, isomorphisms preserve the order of the elements.) In particular, the group $\mathbb{Z}_2\times\mathbb{Z}_2$ is not cyclic.

3. The group $\mathbb{Z}_2\times\mathbb{Z}_3$ has six elements, and its Cayley table is

	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{0},\overline{2})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$	$(\overline{1},\overline{2})$
$(\overline{0},\overline{0})$	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{0},\overline{2})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$	$(\overline{1},\overline{2})$
$(\overline{0},\overline{1})$	$(\overline{0},\overline{1})$	$(\overline{0},\overline{2})$	$(\overline{0},\overline{0})$	$(\overline{1},\overline{1})$	$(\overline{1},\overline{2})$	$(\overline{1},\overline{0})$
$(\overline{0},\overline{2})$	$(\overline{0},\overline{2})$	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{1},\overline{2})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$
$(\overline{1},\overline{0})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$	$(\overline{1},\overline{2})$	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{0},\overline{2})$
$(\overline{1},\overline{1})$	$(\overline{1},\overline{1})$	$(\overline{1},\overline{2})$	$(\overline{1},\overline{0})$	$(\overline{0},\overline{1})$	$(\overline{0},\overline{2})$	$(\overline{0},\overline{0})$
$(\overline{1},\overline{2})$	$(\overline{1},\overline{2})$	$(\overline{1},\overline{0})$	$(\overline{1},\overline{1})$	$(\overline{0},\overline{2})$	$(\overline{0},\overline{0})$	$(\overline{0},\overline{1})$.

Note that the elements in the first component of each pair are in $\mathbb{Z}_2$ and the element in the second component in $\mathbb{Z}_3$. So for instance the first and second entry in $(\overline{1},\overline{1})$ are different: the first $\overline{1}$ is shorthand for $[1]_2$ and the second $\overline{1}$ is shorthand for $[1]_3$. We know another abelian group with six elements: $\mathbb{Z}_6$. These two groups *are* isomorphic, and an isomorhism $\mathbb{Z}_6\to\mathbb{Z}_2\times\mathbb{Z}_3$ is given by

$$\begin{array}{lll} \overline{0}\longmapsto(\overline{0},\overline{0}) & \overline{2}\longmapsto(\overline{0},\overline{2}) & \overline{4}\longmapsto(\overline{0},\overline{1}) \\ \overline{1}\longmapsto(\overline{1},\overline{1}) & \overline{3}\longmapsto(\overline{1},\overline{0}) & \overline{5}\longmapsto(\overline{1},\overline{2}). \end{array}$$

More on this is in Section 3.4.

Proposition 3.20. *If $G'\subseteq G$ and $H'\subseteq H$ are subgroups, then $G'\times H'$ is a subgroup of $G\times H$.*

Proof. Exercise. □

Remarks 3.21. • Not all subgroups of $G\times H$ are of the form $G'\times H'$ for suitable subgroups $G'\subseteq G$ and $H'\subseteq H$. The examples above contain implicit examples of this.

- If G and H are abelian groups, then $G \times H$ is also abelian:
$$(g,h)*(g',h') = (g *_G g', h *_H h') = (g' *_G g, h' *_H h) = (g',h')*(g,h).$$

One can generalize the construction above to an arbitrary number of groups. Recall that the Cartesian product of the sets $S_1, S_2, \dots, S_n$ is the set

$$S_1 \times S_2 \times \ldots \times S_n := \{(a_1, a_2, \dots, a_n) \mid a_i \in S_i \text{ for } i = 1, 2, \dots, n\}.$$

The Cartesian product is also denoted $\prod_{i=1}^n S_i$.

Definition 3.22. Given n groups $(G_1, *_1), (G_2, *_2), \dots, (G_n, *_n)$, we define a group structure on the Cartesian product $G_1 \times G_2 \times \cdots \times G_n$, by setting

$$(g_1, g_2, \dots, g_n) * (x_1, x_2, \dots, x_n) := (g_1 *_1 x_1, g_2 *_2 x_2, \dots, g_n *_n x_n)$$

for all $(g_1, g_2, \dots, g_n), (x_1, x_2, \dots, x_n) \in G_1 \times G_2 \times \cdots \times G_n$. The group $(G_1 \times \cdots \times G_n, *)$ is called the ***direct product of*** $(G_1, *_1), (G_2, *_2), \dots, (G_n, *_n)$.

Proposition 3.23. *The definition above does indeed yield a group.*

Proof. The proof is very similar to the one for the direct product of two groups. The defining properties of a group are satisfied because they are satisfied on each component. □

Notation 3.24. The direct product of a group by itself n times is denoted

$$G^n := \underbrace{G \times G \times \cdots \times G}_{n \text{ times}}.$$

Examples 3.25. 1. In $\mathbb{Z}^4$, we have $(1,3,6,10) + (0,-2,5,-12) = (1,1,11,-2)$.

2. The map

$$\begin{aligned} \mathbb{Z}^3 &\longrightarrow \mathbb{Z} \times \mathbb{Z}_3 \times \mathbb{Z} \times \mathbb{Z} \\ (a,b,c) &\longmapsto (a, [b]_3, 0, 4c) \end{aligned}$$

is a group homomorphism. Exercise.

3.3 Small finite groups

In this section we classify the smallest groups up to isomorphism. The ***order*** of a group G is the cardinality of G.

Remark 3.26. We know by Theorem 3.2 that if G is a finite group with $\#G = n$, then G is isomorphic to a suitable subgroup of S_n. Therefore, in order to study what the possible groups with n elements look like, a naive approach would be to classify the subgroups of n elements of S_n. In practice this is too hard, because

$$\#S_n = n!,$$

and $n!$ grows too fast with n. Already for $n = 8$ we would need to inspect S_8, which has 40320 elements.

Remark 3.27. Recall part (iii) of Corollary 2.29: if G is a finite group and $\#G$ is a prime number, then G is cyclic. More precisely, let a be a generator of G so that $G = \{a^i \mid i = 0, \dots, p-1\}$, by Proposition 1.65. Then the map

$$\begin{aligned} G &\longrightarrow \mathbb{Z}_p \\ a^i &\longmapsto \bar{i} \end{aligned}$$

is an isomorphism.

3.3.1 Groups of order 1

There is only one group of order 1, up to isomorphism.

More precisely, if $G = \{a\}$, there is only one possible operation $*$ defined on G, and it gives $a * a = a$. The Cayley table of G is

$*$	a
a	a.

One may check that $(G, *)$ is a group. All groups consisting of one element are isomorphic to each other.

3.3.2 Groups of order 2, 3, 5 or 7

As special cases of Remark 3.27, we know that:

- There is only one group of order 2, up to isomorphism: every group of order 2 is isomorphic to $(\mathbb{Z}_2, +)$.
- There is only one group of order 3, up to isomorphism: every group of order 3 is isomorphic to $(\mathbb{Z}_3, +)$.
- There is only one group of order 5, up to isomorphism: every group of order 5 is isomorphic to $(\mathbb{Z}_5, +)$.
- There is only one group of order 7, up to isomorphism: every group of order 7 is isomorphic to $(\mathbb{Z}_7, +)$.

Let us inspect the case of order 2 explicitly. Let $G = \{a, b\}$, with $a \neq b$, and denote the operation by $*$. One of a or b has to be the identity element of G, say it is a (or else swap the roles of a and b). This tells us that at least part of the Cayley table of G has to look like

$*$	a	b
a	a	b
b	b.	

Moreover, b needs to have an inverse, that is, an element that multiplied on both sides with b gives the identity a, and the only option left for b^{-1} is b itself:

$*$	a	b
a	a	b
b	b	a.

You are invited to follow similar steps to determine the only possible Cayley table for a group with three elements.

3.3.3 Groups of order 4

In the second item of Examples 3.19 we saw two groups of order 4: they are $\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$. These two groups are not isomorphic, for instance because $\mathbb{Z}_4$ is cyclic and $\mathbb{Z}_2 \times \mathbb{Z}_2$ is not. We claim that any group of order 4 is isomorphic to either $\mathbb{Z}_4$ or $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Let $\#G = 4$. If G is cyclic, it is isomorphic to $\mathbb{Z}_4$.

Proposition 3.28. *Let $\#G = 4$. If G is not cyclic, then $G \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.*

Proof. Let e denote the identity of G. By item (ii) of Corollary 2.29, the order of any element of G has to be a divisor of 4, but 4 is impossible, or else G would be cyclic. Hence the possible orders are only 1 and 2. This implies that $g^2 = e$, for all $g \in G$. As we saw in the first homework sheet, this implies that G is abelian.

Let x and y be elements in G different from e, and such that $x \neq y$. Then we necessarily have

$$G = \{e, x, y, xy\},$$

namely $xy \notin \{e,x,y\}$. Indeed, it cannot be $xy = x$, or else, by multiplying by x^{-1}, we would get $y = e$; similarly, it cannot be $xy = y$, or else $x = e$; lastly, if we had $xy = e$, by multiplying by x, and since $x^2 = e$, we would get $y = e$. So we may start to fill in the Cayley table, with the information that e is the identity and the square of each element is the identity:

	e	x	y	xy
e	e	x	y	xy
x	x	e		
y	y		e	
xy	xy			$e.$

The remaining spots are easily filled, using again the fact that each square is the identity and the abelianity of G:

	e	x	y	xy
e	e	x	y	xy
x	x	e	xy	y
y	y	xy	e	x
xy	xy	y	x	$e.$

Lastly, we define

$$\begin{aligned} G &\longrightarrow \mathbb{Z}_2 \times \mathbb{Z}_2 \\ e &\longmapsto (\overline{0}, \overline{0}) \\ x &\longmapsto (\overline{0}, \overline{1}) \\ y &\longmapsto (\overline{1}, \overline{0}) \\ xy &\longmapsto (\overline{1}, \overline{1}). \end{aligned}$$

You may check that this is an isomorphism. □

Definition 3.29. The group $\mathbb{Z}_2 \times \mathbb{Z}_2$ is called the ***Klein four-group***, or just the ***Klein group***.

3.3.4 Groups of order 6

We already know that if a group of order 6 is cyclic, it is isomorphic to $\mathbb{Z}_6$. We saw another group of order 6 in the third item of Examples 3.19, namely $\mathbb{Z}_2 \times \mathbb{Z}_3$, but we also saw there that this group *is* cyclic, so that $\mathbb{Z}_2 \times \mathbb{Z}_3 \cong \mathbb{Z}_6$.

Proposition 3.30. *Let $\#G = 6$. If G is not cyclic, then $G \cong S_3$.*

Proof. By item (ii) of Corollary 2.29, the order of any element of G has to be a divisor of 6, and 6 itself is not possible, or else the group G would be cyclic. The only options are then order 1 (for the identity e), 2 or 3.

Let us show that there is at least one element of order 3. If there were no such element, then the square of each elements would be equal to the identity e, and this implies that

G is abelian. But then we could follow the same steps as in the proof of Proposition 3.28, by constructing a subgroup $\{e,x,y,xy\}$ of G with exactly 4 elements, and this contradicts Corollary 2.29, since 4 does not divide 6. Hence there *is* at least one element in G of order 3, call it x.

Since $\pi(x) = 3$, the cyclic subgroup generated by x is $\langle x\rangle = \{e,x,x^2\}$. By Proposition 2.40 (with a full proof in homework sheet 3), we know that $\langle x\rangle$ is a normal subgroup of G, since $\frac{\#G}{\#\langle x\rangle} = \frac{6}{3} = 2$.

Moreover, you may prove as an exercise that since 6 is an even integer, G contains at least one element of order 2, call it y. This in particular means that $y^{-1} = y$. And $y \notin \{e,x,x^2\}$, reasoning on the order of the elements. By Lagrange's theorem G splits in two cosets:

$$\begin{aligned} G &= \{e, x, x^2\} \cup y\{e, x, x^2\} \\ &= \{e, x, x^2, y, yx, yx^2\}. \end{aligned}$$

By the normality of $\langle x\rangle$ and the normality criterion, we know that $y^{-1}xy \in \langle x\rangle = \{e,x,x^2\}$, and we need to figure out which of the three elements yxy is equal to. It cannot be that $yxy = e$, or else, by multiplying on the left and the right by y, we would get $x = e$, a contradiction; and it cannot be that $yxy = x$, or else we would get that $xy = yx$, and you may verify that this leads to $\pi(xy) = 6$, another contradiction. Therefore $yxy = x^2$. By multiplying on the left with y, this implies that $xy = yx^2$.

We are then ready to write down the Cayley table of G:

	e	x	x^2	y	yx	yx^2
e	e	x	x^2	y	yx	yx^2
x	x	x^2	e	yx^2	y	yx
x^2	x^2	e	x	yx	yx^2	y
y	y	yx	yx^2	e	x	x^2
yx	yx	yx^2	y	x^2	e	x
yx^2	yx^2	y	yx	x	x^2	$e.$

Now take the cycles $\sigma := (1,2,3)$ and $\tau := (1,2)$ in S_3. You may check that

$$S_3 = \{\mathrm{id}, \sigma, \sigma^2, \tau, \tau\sigma, \tau\sigma^2\},$$

and that the map given by $x \mapsto \sigma$ and $y \mapsto \tau$ defines an isomorphism $G \to S_3$. □

Example 3.31. In the first homework sheet, you saw that the group $\mathrm{GL}_2(\mathbb{Z}_2)$ of invertible 2×2 matrices with coefficients in $\mathbb{Z}_2$ consists of six elements, and that it is not abelian (and hence in particular not cyclic). The proposition above implies that $\mathrm{GL}_2(\mathbb{Z}_2) \cong S_3$.

3.3.5 Groups of order 8

There are five groups of order 8, up to isomorphism:

$$\text{abelian} : \begin{cases} \mathbb{Z}_8 \\ \mathbb{Z}_4 \times \mathbb{Z}_2 \\ \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \end{cases} \qquad \text{non-abelian} : \begin{cases} D_4 & \text{(dihedral group of a square)} \\ Q & \text{(quaternion group).} \end{cases}$$

The group Q consists of the elements 1 (the identity), i, j and k, and all possible products of these elements, with the condition that

$$i^2 = j^2 = k^2 = ijk.$$

This group is a sort of generalization of the complex numbers, and one typically denotes the element ijk by -1 exactly to mimic the behavior of the complex number i. We do not provide any additional details on the two non-abelian groups, or why the list ends here.

For what concerns the three possible abelian groups, this is the topic of the following subsection. Observe that the three given groups are not isomorphic to each other, for instance because $\mathbb{Z}_8$ is the only one that contains an element of order 8, and $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ is the only one that does not contain elements of order 4.

3.4 Finitely generated abelian groups

Recall from Definition 1.56 that a group G is *finitely generated* if there is a finite set of generators for G, that is, if there exist $g_1, \dots, g_m \in G$ such that $G = \langle g_1, \dots, g_m \rangle$. In this section we focus on groups that are finitely generated and abelian.

Examples 3.32.

- A special kind of finitely generated abelian groups is given by cyclic groups: $\mathbb{Z}$ and $\mathbb{Z}_n$, for all $n \in \mathbb{Z}_{>0}$, are cyclic, hence finitely generated and abelian.
- Consider the direct product of $\mathbb{Z}$ with itself n times, that is, the set

$$\mathbb{Z}^n := \underbrace{\mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}}_{n \text{ times}}$$

equipped with component-wise sum. (This is a special case of Definition 3.22.) If $n > 1$, then $\mathbb{Z}^n$ is not cyclic, but it is still abelian and finitely generated: consider the "vectors"

$$e_1 := (1,0,0,\dots,0,0), \qquad e_2 := (0,1,0,\dots,0,0), \qquad \dots, \qquad e_n := (0,0,0,\dots,0,1)$$

in $\mathbb{Z}^n$, that is, each e_i is filled with zeros except for a 1 in the i-th entry. These are generators for $\mathbb{Z}^n$, written $\mathbb{Z}^n = \langle e_1, e_2, \dots, e_n \rangle$, because for every "vector" $(a_1, a_2, \dots, a_n) \in \mathbb{Z}^n$ we may write $(a_1, a_2, \dots, a_n) = a_1 e_1 + a_2 e_2 + \cdots + a_n e_n$.

- The same as in the previous item is true if we replace any of the factors by $\mathbb{Z}_{n_i}$, for some $n_i \in \mathbb{Z}_{>0}$. More generally, the direct product of finitely generated abelian groups is finitely generated and abelian.

In the following, use the same notation as above for $e_1, \dots, e_n$.

Proposition 3.33 (Universal Property of $\mathbb{Z}^n$)**.** *Let G be an abelian group, and $x_1, \dots, x_n \in G$. There is a unique group homomorphism $\varphi \colon \mathbb{Z}^n \to G$ with $\varphi(e_i) = x_i$ for all $i \in \{1, \dots, n\}$.*

Proof. First of all we show the existence: for all $(a_1, \dots, a_n) \in \mathbb{Z}^n$, define

$$\varphi((a_1, \dots, a_n)) := a_1 x_1 + \cdots + a_n x_n.$$

(In the right-hand side, recall that for instance $a_1 x_1$ stands for the sum $x_1 + \cdots + x_1$ with a_1 terms.) This is a homomorphism: for all $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ in $\mathbb{Z}^n$, we have

$$\begin{aligned}
\varphi(a+b) &= \varphi((a_1 + b_1,\, a_2 + b_2, \dots,\, a_n + b_n)) \\
&= (a_1 + b_1)x_1 + (a_2 + b_2)x_2 + \cdots + (a_n + b_n)x_n \\
&= a_1 x_1 + \cdots + a_n x_n + b_1 x_1 + \cdots + b_n x_n \\
&= \varphi(a) + \varphi(b),
\end{aligned}$$

where we could rearrange the terms as G is abelian.

Next, we show the uniqueness. That is, we show that if $\psi\colon \mathbb{Z}^n \to G$ is a homomorphism that satisfies $\psi(e_i) = x_i$ for all i, then $\psi = \varphi$. For all $a = (a_1, \dots, a_n) \in \mathbb{Z}^n$, we have

$$\begin{aligned}\psi(a) &= \psi(a_1 e_1 + \dots + a_n e_n)\\ &\overset{(*)}{=} \psi(a_1 e_1) + \dots + \psi(a_n e_n)\\ &\overset{(**)}{=} a_1 \psi(e_1) + \dots + a_n \psi(e_n)\\ &= a_1 x_1 + \dots + a_n x_n\\ &= \varphi(a),\end{aligned}$$

where in equalities $(*)$ and $(**)$ we used the fact that ψ is a homomorphism. □

Remark 3.34. With the notation of the previous proposition, $\mathrm{im}(\varphi) = \langle x_1, \dots, x_n \rangle$.

Corollary 3.35. *Every finitely generated abelian group is a quotient of $\mathbb{Z}^n$, for a suitable n.*

Proof. Let G be a finitely generated abelian group, and say that G is generated by n elements $x_1, \dots, x_n$. By the previous proposition there is a unique homomorphism $\varphi\colon \mathbb{Z}^n \to G$ with $\varphi(e_i) = x_i$, and this homomorphism is surjective. This means that

$$\mathbb{Z}^n / \ker(\varphi) \cong G,$$

by the first isomorphism theorem. □

3.4.1 Structure theorem for finitely generated abelian groups

By Corollary 3.35, if we wish to classify all possible finitely generated abelian groups up to isomorphism, the task is equivalent to that of understanding the quotients of $\mathbb{Z}^n$, for abritrary n.

Recall that two integers are *coprime* if their greatest common divisor is 1.

Theorem 3.36 (Chinese Remainder Theorem). *Let $m_1, m_2, \dots, m_t$ be pairwise coprime integers, and denote $M := m_1 \cdot m_2 \cdot \ldots \cdot m_t$. Then*

$$\mathbb{Z}_M \cong \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \dots \times \mathbb{Z}_{m_t}.$$

Proof idea. One may show that the map

$$[x]_M \longmapsto \big([x]_{m_1}, [x]_{m_2}, \dots, [x]_{m_t}\big)$$

is an isomorphism. □

Example 3.37. For instance we already observed in the previous section that $\mathbb{Z}_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_3$. You may have implicitly used this theorem when solving systems of congruences. In simpler terms, this theorem means that

$$x \equiv y \mod 6 \qquad \Leftrightarrow \qquad \begin{cases} x \equiv y \mod 2 \\ x \equiv y \mod 3. \end{cases}$$

Theorem 3.38 (Structure Theorem). *Any finitely generated abelian group is isomorphic to a group of the form*

$$\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \mathbb{Z}_{n_s} \times \mathbb{Z}^c,$$

for $n_1, \dots, n_s, c \in \mathbb{N}$, with n_i being a divisor of n_{i+1} for all $i \in \{1, \dots, s-1\}$. This representation is unique.

Proof idea. A formal proof would be too long. Essentially, one may build a machinery similar to that of Gaussian Elimination for matrices. In our context we need to consider matrices with entries in $\mathbb{Z}$. These are more complicated than matrices over $\mathbb{R}$ or $\mathbb{C}$, since unlike $\mathbb{R}$ or $\mathbb{C}$, in $\mathbb{Z}$ we usually cannot perform division. One may still define "elementary matrices", as in the case of Gaussian Elimination, and multiplying a fixed matrix A on the left or the right by an elementary matrix corresponds to performing operations on the rows or columns of A. By performing such operations, one may eventually rewrite a group isomorphically in the form given in the statement. □

Remark 3.39. In the cyclic groups $\mathbb{Z}_n$, all elements have finite order. In $\mathbb{Z}$, all elements except for 0 have infinite order. In a general abelian group, there will be some elements with finite order and some with infinite order. The factors $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \mathbb{Z}_{n_s}$ in the structure theorem are intuitively containing the elements of finite order. When they are mixed with the last factor $\mathbb{Z}^c$, one gets the elements of infinite order.

Example 3.40.

- For instance $\mathbb{Z}_6$ is already given as in the statement of the structure theorem. By the Chinese remainder theorem, one has $\mathbb{Z}_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_3$, but this is not the form prescribed by the structure theorem.

- Say we are given
$$G := \mathbb{Z}_{2^4} \times \mathbb{Z}_{2^2} \times \mathbb{Z}_2 \times \mathbb{Z}_{3^2} \times \mathbb{Z}_{5^2} \times \mathbb{Z}_{5^2}.$$
Even if we permute the order of the factors, this representation of G is not as in the statement of the structure theorem. Instead, we may "group" some factors by noting that
$$\mathbb{Z}_{2^4} \times \mathbb{Z}_{3^2} \times \mathbb{Z}_{5^2} \cong \mathbb{Z}_{2^4.3^2.5^2} \quad \text{and} \quad \mathbb{Z}_{2^2} \times \mathbb{Z}_{5^2} \cong \mathbb{Z}_{2^2.5^2}$$
by the Chinese remainder theorem. So then we may write
$$G \cong \mathbb{Z}_2 \times \mathbb{Z}_{2^2.5^2} \times \mathbb{Z}_{2^4.3^2.5^2},$$
and this *is* the representation of G as in the structure theorem.

- For any prime p, one may wish to find all abelian groups of order p^m, for some m. This problem is related to partitioning m, that is, writing m as a sum of positive integeres. For instance, for $m = 4$ there are five partitions:
$$4 = 4, \qquad 4 = 2 + 2, \qquad 4 = 1 + 3, \qquad 4 = 1 + 1 + 2, \qquad 4 = 1 + 1 + 1 + 1.$$
The five pairwise non-isomorphic abelian groups of order p^4 are
$$\mathbb{Z}_{p^4}, \qquad \mathbb{Z}_{p^2} \times \mathbb{Z}_{p^2}, \qquad \mathbb{Z}_p \times \mathbb{Z}_{p^3}, \qquad \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_{p^2}, \qquad \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_p.$$

- More generally, if $n = p_1^{\alpha_1} \cdots p_s^{\alpha_s}$ is a decomposition in powers of pairwise distinct primes, the number of pairwise non-isomorphic abelian groups of order n is the product $\prod_{i=1}^{s} \#\{\text{partitions of } \alpha_i\}$. For instance, we may factor $144 = 3^2 \cdot 2^4$, and the pairwise non-isomorphic abelian groups of order 144 are given in the table:

	$4 = 4$	$3 = 2 + 2$	$4 = 1 + 3$	$4 = 1 + 1 + 2$	$4 = 1 + 1 + 1 + 1$
$2 = 2$	$\mathbb{Z}_{3^2} \times \mathbb{Z}_{2^4}$	$\mathbb{Z}_{3^2} \times \mathbb{Z}_{2^2} \times \mathbb{Z}_{2^2}$	$\mathbb{Z}_{3^2} \times \mathbb{Z}_2 \times \mathbb{Z}_{2^3}$	$\mathbb{Z}_{3^2} \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{2^2}$	$\mathbb{Z}_{3^2} \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$
$2 = 1 + 1$	$\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_{2^4}$	$\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_{2^2} \times \mathbb{Z}_{2^2}$	$\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_2 \times \mathbb{Z}_{2^3}$	$\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_{2^2}$	$\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$

Chapter 4

Rings

So far we have studied sets with one binary operation. However, the well-known sets of integer, rational, real and complex numbers come with two familiar operations, namely addition and multiplication. From now on, we will consider sets with two binary operations. On one hand, this is more intuitive. On the other hand, considering two binary operations also adds complexity to the study of algebraic structures.

4.1 Definition and basic properties

Definition 4.1. A ***ring*** $(R,+,\cdot)$ consists of a set R with two binary operations $+$ and $\cdot$, which we call ***addition*** and ***multiplication***, satisfying the following conditions:

1. $(R,+)$ is an abelian group;

2. $a\cdot(b\cdot c)=(a\cdot b)\cdot c$ for all $a,b,c\in R$ (***associativity of multiplication***);

3. for all $a,b,c\in R$, we have
$$a\cdot(b+c)=(a\cdot b)+(a\cdot c) \qquad \text{and} \qquad (a+b)\cdot c=(a\cdot c)+(b\cdot c)$$
(***left distributive law*** and ***right distributive law***, respectively).

If additionally the multiplication is commutative, i.e., $a\cdot b=b\cdot a$ for all $a,b\in R$, then R is called a ***commutative ring***.

Remarks 4.2.
- We commonly denote multiplication simply by juxtaposition, writing ab instead of $a\cdot b$.

- The usual convention is that multiplication is performed before addition in the absence of parentheses, e.g.,
$$ab+ac=(a\cdot b)+(a\cdot c).$$

- As for groups, sometimes we will refer to a "ring R", instead of a ring $(R,+,\cdot)$, with the understanding that there are underlying binary operations $+$ and $\cdot$.

- Occasionally we refer to $(R,+)$ as the ***additive group*** of the ring $(R,+,\cdot)$.

- The additive identity of R is denoted 0. The additive inverse of $a\in R$ is denoted $-a$. They are unique.

- $(R,\cdot)$ is a semigroup (see the footnote to Definition 1.16).

Definition 4.3. Let $(R,+,\cdot)$ be a ring. If there is an element $1 \in R$, called ***multiplicative identity*** or ***unity*** of R, such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in R$, then we call R a ***unital ring*** or a ***ring with unity***.

Remarks 4.4.

- The most important rings in this course will be unital rings.
- The unity in a unital ring is unique.
- In some textbooks (and for instance in the past implementations of this course), the existence of a multiplicative identity is required in the definition of a ring, and then "ring" means unital ring.
- A priori, one might have $0 = 1$, so in some definitions, such as that of a field, we will explicitly require that $0 \neq 1$. The (unital!) ring $\{0\}$ is called the ***zero ring***.
- If R is a unital ring, then $(R,\cdot)$ is a monoid (see the footnote to Definition 1.16).

Examples 4.5.

1. $(\mathbb{Z},+,\cdot)$, $(\mathbb{Q},+,\cdot)$, $(\mathbb{R},+,\cdot)$, and $(\mathbb{C},+,\cdot)$ are commutative rings with unity. From now on, when we write simply $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$, we will be referring to these rings. The additive identity is 0 and the unity is 1.

2. $(2\mathbb{Z},+,\cdot)$ is a commutative ring but not a unital ring.

3. Let R be a unital ring and let n be a positive integer. We denote by $M_n(R)$ the set of all $n \times n$ matrices with entries in R. The addition and multiplication in R allows us to add and multiply matrices in the usual manner. Then $M_n(R)$ together with matrix addition and matrix multiplication is a unital ring, with the additive identity given by the zero matrix and the unity given by the matrix with ones on the main diagonal and zeros elsewhere else. In particular, $M_n(\mathbb{Z})$, $M_n(\mathbb{Q})$, $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ are unital rings. They are not commutative rings, for $n \geq 2$.

4. $(\mathbb{N},+,\cdot)$ is not a ring, since $(\mathbb{N},+)$ is not a group. However, the other two conditions in the definition of a ring are satisfied.

5. Let X be any set and let R be any ring. Then the set $R^X := \{\text{functions } X \to R\}$ becomes a ring with addition and multiplication defined pointwise: for all $x \in X$,

$$(f+g)(x) := f(x) + g(x) \quad \text{and} \quad (f \cdot g)(x) := f(x) \cdot g(x).$$

The zero function $f(x) = 0$ for all $x \in X$ plays the role of the additive identity. If R is unital, the function defined by $f(x) = 1_R$ for all $x \in X$ plays the role of the unity of R^X.

6. The set $\mathbb{Z}_n$ under addition and multiplication modulo n forms a commutative ring with unity. The additive identity is $\overline{0}$ and the unity is $\overline{1}$.

7. Let R be a commutative ring with unity. The set of polynomials with coefficients in R

$$R[x] := \{a_0 + a_1 x + \ldots + a_n x^n \mid n \in \mathbb{N},\, a_k \in R \text{ for } k = 0,\ldots,n\}$$

is a commutative ring with unity, under the usual addition and multiplication of polynomials.

8. If $R_1, R_2, \ldots, R_n$ are rings, then $R_1 \times R_2 \times \ldots \times R_n$ together with componentwise addition and multiplication is a ring. The ring axioms for $R_1, R_2, \ldots, R_n$ follow from the ring axioms for each of the components.

9. Let S be a set. Its power set $\mathcal{P}(S)$ forms a unital ring with addition defined as symmetric difference
$$A \triangle B := (A \backslash B) \cup (B \backslash A)$$
and multiplication defined as $A \cdot B := A \cap B$. The additive identity is $\emptyset$, and the unity is S. The additive inverse of A is A itself.

Definition 4.6. An element u of a unital ring R is called a ***unit*** if u is multiplicatively invertible, i.e., there exists $v \in R$ such that $uv = vu = 1$. We denote $R^* := \{\text{units of } R\}$.

It is important not to confuse the notions of *unity* and *unit*. While the unity (i.e., multiplicative identity) is a unit, the converse is in general not true.

Examples 4.7. We have

$$\begin{aligned} \mathbb{Q}^* &= \mathbb{Q} \setminus \{0\} & \mathbb{Z}_n^* &= \{\overline{a} \in \mathbb{Z}_n \mid \gcd(a,n) = 1\} \\ \mathbb{R}^* &= \mathbb{R} \setminus \{0\} & \mathbb{Z}^* &= \{-1, +1\} \\ \mathbb{C}^* &= \mathbb{C} \setminus \{0\} & M_n(\mathbb{R})^* &= \mathrm{GL}_n(\mathbb{R}). \end{aligned}$$

Lemma 4.8. *Let R be a ring. Then, for all positive integers m and n, and for all $a, a_1, \ldots, a_m$, $b, b_1, \ldots, b_n \in R$,*

$$\begin{aligned} a(b_1 + \ldots + b_n) &= ab_1 + \ldots + ab_n \\ (a_1 + \ldots + a_m)b &= a_1 b + \ldots + a_m b \\ (a_1 + \ldots + a_m)(b_1 + \ldots + b_n) &= a_1 b_1 + a_1 b_2 + \ldots + a_m b_n. \end{aligned}$$

Proof. Exercise. □

Example 4.9. If R is a ring and $a, b \in R$, then

$$(a+b)^2 = (a+b)(a+b) = a(a+b) + b(a+b) = a^2 + ab + ba + b^2.$$

Unless we have a commutative ring, this is in general not equal to $a^2 + 2ab + b^2$. However, in $\mathbb{Z}_2$, we have even $(a+b)^2 = a^2 + b^2$.

Since condition (1) in the definition of a ring concerns only addiiton, and the conditions (2) and (3) concern only multiplication, note that in order to prove a statement involving both operations, we need to use the distributive law.

Lemma 4.10. *Let R be a ring and $a, b \in R$. Then*

1. $a0 = 0a = 0$,
2. $a(-b) = (-a)b = -(ab)$.

Proof.
1. We have $0 + a0 = a0 = a(0+0) = a0 + a0$. By the cancellation law in $(R, +)$, we have $0 = a0$. The identity $0a = 0$ can be shown using an analogous proof.
2. We have $ab + a(-b) = a(b + (-b)) = a0 = 0$, where the last equality follows by part (1). Hence $a(-b)$ is the additive inverse of ab, i.e., $-(ab) = a(-b)$. The identity $-(ab) = (-a)b$ can be shown using an analogous proof.

□

Remarks 4.11.
- Since $(-a)b = -(ab)$ by Lemma 4.10 part (2), the we can write $-ab$ without an ambiguity.

- If $R \neq \{0\}$ is a unital ring, then $1 \neq 0$. We can show the converse statement that if $1 = 0$, then $R = \{0\}$. Suppose $1 = 0$, then $a = a1 = a0 = 0$ for all $a \in R$.

Recall from Definition 1.41 that in group theory we denote

$$na := \underbrace{a + a + \cdots + a}_{n \text{ terms}}.$$

This is not the multiplication in the ring, as n might not be a ring element at all. It is only an abbreviation of $a + \cdots + a$. In case $n < 0$, then

$$na := \underbrace{(-a) + (-a) + \ldots + (-a)}_{-n \text{ terms}}.$$

Finally, we set

$$0_{\mathbb{Z}} a := 0_R.$$

Then, for all $m, n \in \mathbb{Z}$ and for all $a, b \in R$, this "multiplication by scalars in $\mathbb{Z}$, satisfies

$$(m+n)a = ma + na, \qquad (mn)a = m(na), \qquad n(a+b) = na + nb.$$

It's a good exercise to go through these identities and make sense of every operation involved. If for instance we denote the "multiplication by integer scalars" we just defined with just a dot

$$\cdot \colon \mathbb{Z} \times R \to R,$$

and the respective sum and multiplication of $\mathbb{Z}$ and R with their corresponding indeces, then the three identities above can be written as

$$(m +_{\mathbb{Z}} n) \cdot a = m \cdot a +_R n \cdot a, \qquad (m \cdot_{\mathbb{Z}} n) \cdot a = m \cdot (n \cdot a), \qquad n \cdot (a +_R b) = n \cdot a +_R n \cdot b.$$

Lemma 4.12. *Let R be a unital ring. For all $a, b \in R$ and for all $m, n \in \mathbb{Z}$,*

1. $n \cdot a = (n \cdot 1)a = a(n \cdot 1)$,
2. $n \cdot (ab) = (n \cdot a)b = a(n \cdot b)$,
3. $(m \cdot a)(n \cdot b) = (mn) \cdot (ab)$.

Proof. 1. If $n = 0$, then all the products are 0. Let $n > 0$. Then

$$(n \cdot 1)a = (1 + \ldots + 1)a = a + \ldots + a = n \cdot a.$$

Recall that $(-n) \cdot 1 = n \cdot (-1)$ and $(-n) \cdot a = n \cdot (-a)$ by definition. Hence

$$((-n) \cdot 1)a = (n \cdot (-1))a = ((-1) + \ldots + (-1))a = (-a) + \ldots + (-a) = n \cdot (-a) = (-n) \cdot a.$$

The third equality follows by distributivity and Lemma 4.10 part (b): $(-1)a = -(1a) = -a$.

2. We have

$$n \cdot (ab) = (n \cdot 1)(ab) = ((n \cdot 1)a)b = (n \cdot a)b.$$

The first and third equality follow by the previous item and the second one by associativity of multiplication.

3. We have

$$(m \cdot a)(n \cdot b) = m \cdot (a(n \cdot b)) = m \cdot (n \cdot (ab)) = (mn) \cdot (ab).$$

The first two equalities follow by the previous item and the last one by the identity from group theory.

□

4.2 Some important types of rings

4.2.1 Integral domains

In $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$, the product of two elements is zero if and only if at least one of the factors is zero. This is an important property of our usual number system and we might not realize that we are using it. For example, consider the equation

$$x^2 - 3x + 2 = 0.$$

Factoring the left-hand side gives

$$(x-1)(x-2) = 0.$$

From this we conclude that $x = 1$ or $x = 2$. The reason behind this conclusion is that if the product of two numbers is zero, then one of the two numbers has to be zero.

The same is not true in all rings: In $\mathbb{Z}_{12}$ we have $\overline{3} \cdot \overline{4} = \overline{0}$, and neither factor is $\overline{0}$.

Definition 4.13. Let R be a ring. We say that $a \in R \setminus \{0\}$ is a ***zero-divisor*** if there is an element $b \in R \setminus \{0\}$ such that

$$ab = 0 \quad \text{or} \quad ba = 0.$$

Example 4.14. The matrix $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ is a zero-divisor in $M_2(\mathbb{R})$, since

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 6 & 2 \\ -3 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Lemma 4.15. *Let R be a unital ring. If $a \in R$ has a multiplicative inverse, then a is not a zero-divisor.*

Proof. Suppose $ba = 0$ and c is the multiplicative inverse of a. Then

$$b = b1 = b(ac) = (ba)c = 0c = 0,$$

hence $b = 0$, and a is not a zero-divisor. □

Example 4.16. An element $\overline{a} \in \mathbb{Z}_n$ is a zero-divisor if and only if $\gcd(a,n) \neq 1$. If $\gcd(a,n) \neq 1$, we can take $\overline{b} := \overline{n/\gcd(a,n)}$, and then $\overline{ab} = \overline{0}$.

Definition 4.17. Let R be a nonzero ring. We say that R is a ***domain*** if R has no zero-divisiors. If in addition R is commutative, then we say that R is an ***integral domain***.

Examples 4.18.

1. The rings $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are integral domains.
2. If R is an integral domain, then the polynomial ring $R[x]$ with coefficients in R is an integral domain.
3. The ring $\mathbb{Z}_n$ is an integral domain if and only if n is prime.
4. The ring $M_n(\mathbb{R})$ is not a domain if $n > 1$.

Lemma 4.19 (Cancellation law)**.** *If D is a domain and $a \in D \setminus \{0\}$, then $ab = ac$ implies that $b = c$ for all $b, c \in D$.*

Proof. The equality $ab = ac$ implies $ab - ac = 0$ and $a(b-c) = 0$. Since $a \neq 0$, then $b - c = 0$ and $b = c$. □

Remarks 4.20. • If D is not a domain, then the cancellation law does not hold. Let $a, b \in \mathbb{R}$ such that $a \neq 0$, $b \neq 0$ and $ab = 0$. Then $ab = a0$, but $b \neq 0$. In $\mathbb{Z}_{12}$ we have $\overline{3} \cdot \overline{4} = \overline{3} \cdot \overline{0}$, but it is not true that $\overline{3} = \overline{0}$.

- If D is an integral domain, then one can solve polynomial equations in $D[x]$ by factoring them and setting each factor to zero.

Example 4.21. Consider the equation $x^3 + \overline{3}x = \overline{0}$ in $\mathbb{Z}_7$, which is an integral domain. The left-hand side factors as $x(x^2+\overline{3}) = x(x^2-\overline{4}) = x(x-\overline{2})(x+\overline{2})$. Hence the solutions are $x = \overline{0}, \overline{2}, \overline{5}$.

Definition 4.22. Let R be a ring with unity 1_R. The ***characteristic*** of R is

$$\operatorname{char}(R) := \begin{cases} \min\Big\{n \in \mathbb{Z}_{>0} \mid \underbrace{1_R + \cdots + 1_R}_{n \text{ times}} = 0_R\Big\} & \text{if this minimum exists,} \\ 0 & \text{otherwise.} \end{cases}$$

That is, in case the minimum exists, $\operatorname{char}(R)$ is the group-theoretic order of 1_R in $(R, +)$.

Example 4.23. The characteristic of $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ is 0. The characteristic of $\mathbb{Z}_n$ is n.

Remark 4.24. Let D be a unital domain. The additive order of any nonzero element is equal to that of 1_R: suppose $a \in D \setminus \{0\}$, and $na = 0$. Then $0 = na = n(1_D a) = (n1_D)a$. Since D does not have zero divisors, this is equivalent to $n1_D = 0$.

Lemma 4.25. *The characteristic of a unital integral domain D is either* 0 *or a prime number.*

Proof. Assume that $\operatorname{char}(D) = n \neq 0$. Write $n = n_1 n_2$. Then

$$0 = n1_D = (n_1 n_2)1_D = (n_1 1_D)(n_2 1_D).$$

Since D does not have zero-divisors, either $n_1 1_D = 0$ or $n_2 1_D = 0$. Without loss of generality assume that $n_1 1_D = 0$. But by definition of $\operatorname{char}(D)$, it must be $n_1 = n$. Hence the only way to factor n is $n = n \times 1$, which means that n is a prime number. □

4.2.2 Fields

Definition 4.26. Let F be a unital ring with $1 \neq 0$. If every nonzero element of F is a unit, then F is called a ***division ring*** (or a ***skew field***). A commutative division ring is called a ***field***.

Remark 4.27. A triple $(F, +, \cdot)$ is a field if and only if the following conditions are satisfied:

1. $(F, +)$ is an abelian group (the *additive group* of F),

2. $(F \setminus \{0\}, \cdot)$ is an abelian group (the *multiplicative group* of F), and

3. $a(b+c) = ab + ac$ and $(a+b)c = ac + bc$, for all $a, b, c \in F$.

Examples 4.28. 1. $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$ are fields.

2. $(\mathbb{Z}, +, \cdot)$ is not a division ring (and hence also not a field), since 1 and -1 are the only units.

3. A ***quaternion*** is an expression of the form

$$a + bi + cj + dk,$$

where a, b, c, d are real numbers and i, j, k are symbols satisfying

$$i^2 = j^2 = k^2 = ijk = -1.$$

We have $ij = -ji$, and hence the multiplication of quaternions is not commutative. The set of quaternions with addition and multiplication forms a skew field.

4. The set of all ***rational functions*** over $\mathbb{R}$ is defined as

$$\mathbb{R}(x) := \left\{ \frac{p(x)}{q(x)} \mid p(x), q(x) \in \mathbb{R}[x], q(x) \neq 0 \right\}.$$

The set $\mathbb{R}(x)$ is a field under pointwise addition and multiplication.

Proposition 4.29. *(1) Every field is an integral domain.*

(2) Every finite integral domain with unity is a field.

Proof. Commutativity and having a unity are evident in both statements.

(1) Let F be a field. Let $a \in F$, $a \neq 0$. Then a is a unit and hence not a zero-divisor by Lemma 4.15. Therefore F is an integral domain.

(2) Let D be a finite integral domain. We need to show that every $a \in D \setminus \{0\}$ has a multiplicative inverse. Consider the set $D' := \{ax \mid x \in D\}$. Whenever $x_1 \neq x_2$, the cancellation law (Lemma 4.19) implies that $ax_1 \neq ax_2$. Hence $D' = D$, and $a\hat{x} = 1$ for some $\hat{x} \in D$. Since D is commutative, also $\hat{x}a = 1$, and hence $a^{-1} = \hat{x}$. □

The finiteness assumption is fundamental in part (2). Indeed, $\mathbb{Z}$ is an integral domain but it is not a field.

It follows from Proposition 4.29 that $\mathbb{Z}_n$ is a field exactly when $\mathbb{Z}_n$ is an integral domain (which happens if and only if n is a prime number).

4.3 Subrings and ideals

Definition 4.30. Let R be a ring. A subset $S \subseteq R$ is called a ***subring*** of R if S is a ring under the induced operations from R.

Proposition 4.31 (Subring criterion)**.** *Let $(R, +, \cdot)$ be a ring and $S \subseteq R$. Then S is a subring of R if and only if the following conditions hold:*

(S1) $(S, +)$ is a subgroup of $(R, +)$,

(S2) for all $a, b \in S$, we have $ab \in S$.

Proof. The conditions (S1) and (S2) guarantee that $(S, +, \cdot)$ is a ring, since the associativity and distributivity hold in S simply because they hold in R and we are just restricting the operations. Conversely, by definition of ring, $(S, +)$ has to be a group, so that for all elements $a, b \in S$, we have $a - b \in S$ and the additive identity has to be in S, which mean that $(S, +)$ is a subgroup of $(R, +)$. And moreover for $(S, +, \cdot)$ to be a ring, the restriction of $\cdot$ has to be a binary operation on S, so that (S2) holds, too. □

Note in particular that a subring S has to contain 0_R, and 0_R is the additive identity of S.

Example 4.32. 1. In the chain $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$, each ring is a subring of the next.

2. The set of all continuous functions from $[0,1]$ to $\mathbb{R}$ is a subring of all functions from $[0,1]$ to $\mathbb{R}$. The same is true for differentiable functions.

3. Upper triangular $n \times n$ matrices with entries in a ring R form a subring of $M_n(R)$.

4. A ***Gaussian integer*** is a complex number of the form $a + bi$, for $a, b \in \mathbb{Z}$. The Gaussian integers with ordinary addition and multiplication of complex numbers form a subring of $\mathbb{C}$.

5. Let R be a commutative ring with unity. Then R is a subring of $R[x]$.

6. Fix $n \in \{-1, \pm 2, \pm 3, \ldots\}$. The set

$$\mathbb{Z}[\sqrt{n}] := \{a + b\sqrt{n} \mid a, b, \in \mathbb{Z}\}$$

is a subring of $\mathbb{C}$ and a subring of $\mathbb{R}$ if $n > 0$. If $n = -1$, then we get the Gaussian integers.

Remark 4.33. If one defines rings as unital rings, namely always assuming the existence of a multiplicative identity, then the definition of a subring differs: one additionally requires that the unity in S is the same as the unity in R.

Examples 4.34. 1. In the previous examples, all rings are unital rings, and the unity of the given subrings is the same as the unity of the ambient ring.

2. All matrices of the form $\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$, where $a \in \mathbb{R}$, form a subring of $M_2(\mathbb{R})$. However, note that its unity is $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, which is not the unity of $M_2(\mathbb{R})$.

Lemma 4.35. *Let R be a domain with unity 1_R, and let $S \subseteq R$ be a subring with unity $1_S \neq 0$. Then $1_S = 1_R$.*

Proof. Because both 1_S and 1_R are multiplicative identities, we have $1_S 1_S = 1_S = 1_S 1_R$, which we may rewrite as

$$1_S(1_S - 1_R) = 1_S 1_S - 1_S 1_R = 0.$$

But since R is a domain, there are two options: either $1_S = 0$, which is not the case, or $1_S - 1_R = 0$, which means exactly that $1_S = 1_R$. □

4.3.1 Ideals

In group theory, we first studied subgroups, and then among those we considered the special class of normal subgroups. In ring theory, the corresponding special subrings are not called "normal subrings", but instead *ideals*.

Definition 4.36. Let R be a ring and let $I \subseteq R$ be a subset. Then I is called an ***ideal*** of R if:

(I1) $(I, +)$ is a subgroup of $(R, +)$, and

(I2) for all $r \in R$ and $a \in I$, we have $ra \in I$ and $ar \in I$.

Examples 4.37. • Every ring R has the ideals R and $\{0\}$.

- The subgroups of $(\mathbb{Z},+)$ are exactly the subsets of the form $n\mathbb{Z}$. They are also ideals in the ring $\mathbb{Z}$, since $r(nm) = (nm)r = n(mr) \in n\mathbb{Z}$ for all $r \in \mathbb{Z}$. Since there are no other subgroups of $\mathbb{Z}$, there are no other ideals of $\mathbb{Z}$.

Remarks 4.38.

- An ideal of R is a subring of R.

- Elsewhere in the literature (and for instance in the past implementation of this course), where rings are always assumed to have a unity, subrings are also assumed to have the same unity by definition. In that case, an ideal would in general not be a subring: indeed, by assuming that $1 \in I$, we would get $r1 = r \in I$ for all $r \in R$ and $I = R$. Hence R itself would be the only ideal that is also a subring, in that framework. One of the reasons why we are not assuming all rings to be unital and subrings to share the unity in this course is that we want to stress the analogy

$$\begin{array}{ccc} \{\text{subrings}\} & \longleftrightarrow & \{\text{subgroups}\} \\ \cup & & \cup \\ \{\text{ideals}\} & \longleftrightarrow & \{\text{normal subgroups}\}. \end{array}$$

- Let R be a unital ring. If $I \subsetneq R$ is a proper ideal of R, then $I \cap R^* = \emptyset$. If there were a unit $u \in I$, then $1 = u^{-1}u \in I$ and hence $I = R$.

- A field F has *only* two ideals: F and 0. In fact, one can show that the only commutative unital rings with this property are fields.

Example 4.39. Let $R = \{\text{functions } \mathbb{R} \to \mathbb{R}\}$ and $S = \{f \in R \mid f \text{ is differentiable}\}$. Then S is a subring of R, but not an ideal: for instance, if we take the functions defined as

$$f(x) := 1 \quad \text{and} \quad g(x) := |x| \qquad \text{for all } x \in \mathbb{R},$$

then we have $f \in S$ and $g \in R$, but $fg \notin S$.

Proposition 4.40 (Ideal criterion). *Let R be a ring and $I \subseteq R$. Then I is an ideal if and only if the following holds:*

(1) $I \neq \emptyset$,

(2) for all $a, b \in I$, we have $a - b \in I$,

(3) for all $r \in R$ and $a \in I$, we have $ra \in I$ and $ar \in I$.

Proof. Conditions (1) and (2) are equivalent to (I1) by Proposition 1.54. Condition (3) is exactly the same as (I2). □

Example 4.41. The set $I := \{p \in \mathbb{R}[x] \mid \text{the constant term of } p \text{ is } 0\}$ is an ideal of $\mathbb{R}[x]$. Indeed I is non-empty, since the zero polynomial is in I. Moreover, if we subtract two polynomials with zero constant term, the constant term stays zero. Lastly, if we multiply any polynomial $f = a_n x^n + \cdots + a_1 x + a_0$ by a polynomial $p = b_m x^m + \cdots + b_1 x \in I$, we get

$$fp = (\text{terms of degree} > 1) + a_0 b_1 x,$$

which is in I. Hence I is an ideal.

4.3.2 Generating sets

Recall that in Proposition 1.55 we proved that the intersection of subgroups is a subgroup. That fact was fundamental in order to define the *subgroup generated by a subset*, namely the smallest subgroup containing that subset. In ring theory, it is true that the intersection of subrings is a subring, but we are more interested in the analogous property for ideals, as we want to focus on the *ideal generated by a subset*.

Proposition 4.42 (Intersection of ideals is an ideal). *Let R be a ring and let $\{I_j \mid j \in J\}$ be a family of ideals of R, for some index set J. Then*

$$I := \bigcap_{j\in J} I_j = \{a \in R \mid \textit{for all } j \in J,\ a \in I_j\}$$

is an ideal of R.

Proof. It can be proven using the ideal criterion, similarly to Proposition 1.55. □

Let R be a ring and $S \subseteq R$ a subset. There is at least one ideal of R that contains S, namely R itself. By Proposition 4.42, the intersection of all ideals that contain S is again an ideal. This is the smallest ideal of R containing S.

Definition 4.43. Let R be a ring and $S \subseteq R$ a subset. The smallest ideal of R containing S is called the ***ideal generated by*** S and denoted (S). The elements of S are called the ***generators*** of (S). If $S = \{a_1, \dots, a_k\}$ is finite, then (S) is called ***finitely generated***, and we write

$$(a_1, \dots, a_k) := \big(\{a_1, \dots, a_k\}\big).$$

An ideal of the form (a), namely generated by a single element, is called a ***principal ideal***.[1]

In the group theory, the notion corresponding to a principal ideal is a cyclic subgroup.

Examples 4.44.

- The ideal $\{0\} = (0)$ is a principal ideal.
- Let R be a unital ring. The ideal $R = (1)$ is a principal ideal.
- For an ideal I, we always have $I = (I)$.
- Sometimes an ideal is given with more generators than actually needed. For instance in the ring $\mathbb{Z}$, the ideal $(2,3)$ turns out to be principal, and actually equal to $\mathbb{Z}$.
- To generalize the previous point, for any $a_1, \dots, a_k \in \mathbb{Z}$ one may prove that

  $$(a_1, \dots, a_k) = (\gcd(a_1, \dots, a_k)),$$

  that is, the ideal generated by a bunch of integers is actually the principal ideal generated by their greatest common divisor. This can be shown with the Euclidean algorithm to compute the common divisor: for instance perform the division with remainder

  $$a_1 = qa_2 + r,$$

  where $0 \leq r < a_2$. Then, since $I := (a_1, \dots, a_k)$ is an ideal and $a_1, a_2 \in I$, by we have $r = a_1 - qa_2 \in I$, and in fact

  $$(a_1, a_2, \dots, a_k) = (r, a_2, \dots, a_k).$$

[1] Often in the literature, and for instance in the previous implementations of this course, the ideal generated by a set S is denoted $\langle S \rangle$ instead of (S). In this course we reserve the angle brackets for subgroups.

By performing as many divisions with remainder as necessary, one eventually gets remainders equal to zero, which one may disregard as generators of I. The last remainder is the greatest common divisor of $a_1,\ldots,a_k$.[2]

We describe explicitly the principal ideals in commutative rings:

Proposition 4.45. *Let R be a commutative unital ring and let $a \in R$. The principal ideal generated by a is*

$$(a) = \{ra \mid r \in R\}.$$

Proof. First of all we check that the right-hand side $I := \{ra \mid r \in R\}$ is an ideal, by using the ideal criterion (Proposition 4.40):

1. Indeed $I \neq \emptyset$, because $a = 1a \in I$.
2. Let $x, y \in I$, so that $x = r_1 a$ and $y = r_2 a$, for some $r_1, r_2 \in R$. Then $x - y = (r_1 - r_2)a \in I$.
3. Let $x \in R$ and let $y \in I$, so that $y = ra$ for some $r \in R$. Then $xy = (xr)a \in I$.

So I is an ideal. Next, let J be an arbitrary ideal of R with $a \in J$, and we shall show that $I \subseteq J$. Indeed, since J is an ideal, this is true by condition $(I2)$ in the definition of an ideal. But then I is contained in every ideal containing a, meaning that $I \subseteq (a)$. And clearly $(a) \subseteq I$, so we are done. □

You may compare (a) with the group-theoretic concept of a cyclic subgroup generated by a, that is, $\langle a\rangle = \{ia \mid i \in \mathbb{Z}\}$. The following is immediate, but we state it explicitly nevertheless:

Corollary 4.46. *Let R be a commutative unital ring and let $a \in R$. For any $b \in R$,*

$$b \in (a) \qquad \Leftrightarrow \qquad b = ra \text{ for some } r \in R.$$

Remark 4.47. The union of ideals is in general not an ideal. Consider for instance the ideals (2) and (3) in $\mathbb{Z}$. Their union is not even a subgroup, so it cannot be an ideal.

Definition 4.48. Let R be a commutative ring, and let $I_1, I_2 \subseteq R$ be ideals. The ideal generated by the union $I_1 \cup I_2$ is denoted $I_1 + I_2 := (I_1 \cup I_2)$ and is called the ***sum of*** I_1 ***and*** I_2. In general, for an arbitrary family of ideals $\{I_j \mid j \in J\}$, we denote

$$\sum_{j\in J} I_j := \Big(\bigcup_{j\in J} I_j\Big).$$

Proposition 4.49. *Let R be a commutative unital ring, and let $I_1,\ldots,I_k \subseteq R$ be ideals. Then*

$$I_1 + \cdots + I_k = \{b_1 + \cdots + b_k \mid b_j \in I_j\}.$$

For $a_1,\ldots,a_k \in R$, we have

$$(a_1,\ldots,a_k) = \{r_1a_1 + \ldots + r_ka_k \mid r_i \in R\}.$$

Proof. Exercise. Similar to the proof of Proposition 4.45. □

Examples 4.50. 1. The ideal generated by $x \in \mathbb{R}[x]$ is

$$(x) = \{a_0x + a_1x^2 + \ldots + a_nx^{n+1} \mid n \geq 0, a_i \in \mathbb{R}\}.$$

In fact, this ideal is equal to the one given in Example 4.41.

[2] In number theory, people often use the notation $(a_1,\ldots,a_k)$ to denote the greatest common divisor of $a_1,\ldots,a_k$.

2. All ideals in $\mathbb{Z}$ are principal ideals:

$$(m) = m\mathbb{Z} \quad \text{for all } m \in \mathbb{Z}.$$

Definition 4.51. A ring whose ideals are all principal is called a ***principal ideal ring***. An integral domain whose ideals are all principal is called a ***principal ideal domain*** (abbreviated PID in the literature).[3]

Example 4.52. 1. The ring $\mathbb{Z}$ is a PID. Every ideal is of the form $n\mathbb{Z}$, for $n \in \mathbb{Z}$. This is explained in the last item of Examples 4.44.

2. Any field K is trivially a PID, since its only ideals are (0) and $(1) = K$.

3. The rings $\mathbb{Q}[x]$, $\mathbb{R}[x]$ and $\mathbb{C}[x]$ are PIDs. (More generally, if the coefficients are in a field K, the polynomial ring $K[x]$ is a PID.) This will be proven in Computational Algebraic Geometry. However, the integral domain $\mathbb{Z}[x]$ is not a PID: for example, the ideal $(2,x)$ is not a principal ideal.

4. One may consider polynomial rings with more than one variable. For instance $\mathbb{Q}[x,y]$, $\mathbb{R}[x,y]$ and $\mathbb{C}[x,y]$ are integral domains, but none of them is a PID, because the ideal (x,y) is not principal.

4.3.3 Prime and maximal ideals

Definition 4.53. Let R be a commutative unital ring and I be a proper ideal, that is, $I \subsetneq R$.

- We call I a ***prime ideal*** if the following condition holds: for all $a,b \in R$, the fact that $ab \in I$ implies that $a \in I$ or $b \in I$ (or both).
- We call I a ***maximal ideal*** if for all ideals $J \subseteq R$ containing I, either $J = I$ or $J = R$.

Remark 4.54. In other words:

- A proper ideal is *prime* if when the product of two elements is in the ideal, at least one of the factors has to be in the ideal.
- A proper ideal I is *maximal* if there are no ideals J such that $I \subsetneq J \subsetneq R$. Or still in other words, if I is maximal with respect to inclusion among the proper ideals of R.

Examples 4.55.
- Let K be a field. We know that the only proper ideal is (0). This is both a prime and a maximal ideal of K.
- A ring R is an integral domain if and only if (0) is a prime ideal.

Proposition 4.56. *Let R be a commutative unital ring. If I is maximal ideal, then I is prime.*

Proof. Conversely, we show that if I is not prime, then I is not maximal. Let a and b be elements outside of I satisfying $ab \in I$. Then we show that $R \supsetneq (a) + I$: indeed if we had $R = (a) + I$, then in particular 1_R would be in the right-hand side and could be written as

$$1_R = ac + i,$$

[3] Unlike other sciences, it is not so common in pure mathematics to use abbreviations. In algebra there are very few abbreviations and this is one of them.

for some $c \in R$ and some $i \in I$. Multiplying both sides by b, this would then yield

$$b = abc + ib,$$

and now since $ab \in I$ and $ib \in I$, this would mean that $b \in I$, a contradiction.

Moreover, since $a \notin I$, we clearly have $I \subsetneq (a) + I$. But then we have produced an ideal, $(a) + I$, which is strictly included between I and R, and hence I is not maximal. □

Example 4.57. Consider $\mathbb{Z}$. We know that the ideals of $\mathbb{Z}$ are of the form $(n) = n\mathbb{Z}$, for $n \in \mathbb{N}$. We claim that:

- (n) is a maximal ideal exactly when n is a prime number,
- (n) is a prime ideal exactly when n is a prime number or $n = 0$.

Proof. Let $p \in \mathbb{N}$ be a prime number. Then (p) is a maximal ideal: indeed, by the Euclidean algorithm (and the discussion in the last item of Examples 4.44), if there were some element $a \notin (p)$ and such that $(a) + (p) \subsetneq \mathbb{Z}$, then we would get a contradiction, since

$$(a) + (p) = (a, p) = (\gcd(a, p)),$$

and because p is prime, the greatest common divisor $\gcd(a,p)$ is either 1, in which case $(a, p) = \mathbb{Z}$, or it is equal to p, in which case $a \in (p)$, and both are contradictions. By Proposition 4.56, (p) is then also a prime ideal.

On the other hand, if $n \in \mathbb{N}$ is not a prime number, then we may factor n in a product of primes, say $n = p_1 \cdots p_k$, with $k \geq 2$, and pick $a := p_1$ and $b := p_2 \cdots p_k$. Then definitely $a \notin (n)$ and $b \notin (n)$, but $ab = n \in (n)$. This shows that (n) is not prime, and hence not maximal.

Lastly, (0) is prime ideal: as it happens in any other integral domain, if the product of two integers is 0, at least one of them is 0. □

4.4 Ring homomorphisms

Definition 4.58. Let $(R, +_R, \cdot_R)$ and $(T, +_T, \cdot_T)$ be rings. A map $f\colon R \to T$ is called a ***ring homomorphism*** if the following conditions are satisfied for all $a, b \in R$:

1. $f(a +_R b) = f(a) +_T f(b)$,
2. $f(a \cdot_R b) = f(a) \cdot_T f(b)$.

Remark 4.59.

- The first condition means that a ring homomorphism $f\colon R \to T$ is a group homomorphism from $(R, +_R)$ to $(T, +_T)$. Hence all the results about group homomorphisms hold for ring homomorphisms as well.
- If you like to visualize things pictorially, the fact that f is a ring homomorpism means that both squares in the diagram

$$\begin{tikzcd}
R \arrow[d, "f"'] & R \times R \arrow[l, "+_R"'] \arrow[r, "\cdot_R"] \arrow[d, "f \times f"] & R \arrow[d, "f"] \\
T & T \times T \arrow[l, "+_T"] \arrow[r, "\cdot_T"'] & T
\end{tikzcd}$$

are commutative. That is, for any $(a,b) \in R \times R$ in the top-middle place, you may take either way to the bottom-left corner and end up with the same result, or either way to the bottom-right corner and end up with the same result:

$$\begin{tikzcd}
a +_R b \arrow[d, mapsto] & (a,b) \arrow[l, mapsto] \arrow[r, mapsto] \arrow[d, mapsto] & a \cdot_R b \arrow[d, mapsto] \\
f(a + Rb) = f(a) +_T f(b) & (f(a), f(b)) \arrow[l, mapsto] \arrow[r, mapsto] & f(a) \cdot_T f(b) = f(a \cdot_R b)
\end{tikzcd}$$

- When rings are assumed to be unital rings in the literature, authors usually require an additional condition in the definition above: $f(1_R) = 1_T$.

Examples 4.60.

1. Let $n \in \mathbb{Z}_{>0}$. The map $\varphi : \mathbb{Z} \to \mathbb{Z}_n$, where $\varphi(a)$ is the remainder of a modulo n, is a ring homomorphism. We already know that it is a group homomorphism from $(\mathbb{Z},+)$ to $(\mathbb{Z}_n,+_n)$. Let $a = q_1 n + r_1$ and $b = q_2 n + r_2$. Then

   $$ab = (q_1 n + r_1)(q_2 n + r_2) = (q_1 q_2 n + q_1 r_2 + q_2 r_1)n + r_1 r_2.$$

   Hence $\varphi(ab)$ is the remainder or $r_1 r_2$ modulo n. Also $\varphi(a)\varphi(b)$ is the remainder or $r_1 r_2$ modulo n (it follows from the definition of multiplication in $\mathbb{Z}_n$). Hence φ is a ring homomorphism.

2. Let $F = \mathbb{R}^{\mathbb{R}} = \{\text{functions } \mathbb{R} \to \mathbb{R}\}$, equipped with pointwise sum and multiplication. For any $a \in \mathbb{R}$, the ***evaluation map*** $\varphi_a \colon F \to \mathbb{R}$, given by $\varphi_a := f(a)$ for $f \in F$, is a ring homomorphism.

3. The map

   $$\begin{aligned} \varphi\colon \mathbb{C} &\longrightarrow M_2(\mathbb{R}) && \text{where } I := \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \\ a + ib &\longmapsto \begin{pmatrix}a & b\\ -b & a\end{pmatrix} = aI + bJ, && \text{and } J := \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}, \end{aligned}$$

   is a ring homomorphism. Exercise.

4. Let R be any ring. The identity function $\mathrm{id}_R \colon R \to R$ is a ring homomorphism.

5. Let R be any ring. The zero function $f \colon R \to R$, that is, the function defined as $f(a) = 0$ for all $a \in R$, is a ring homomorphism. This homomorphism does *not* preserve the unity in general, but all other homomorphisms listed above do.

6. Consider the map $\mathbb{Z} \to \mathbb{Z}$ given by $f(n) := 2n$ for all $n \in \mathbb{Z}$. This is a group homomorphism from $(\mathbb{Z},+)$ to $(\mathbb{Z},+)$, but it is not a ring homomorphism, for instance because $f(2)f(3) = 24 \neq 12 = f(2 \cdot 3)$.

Remark 4.61. Since a ring homomorphism $f \colon R \to R'$ is a group homomorphism from $(R,+)$ to $(R',+)$, we immediately have that $f(0_R) = 0'_R$ and $f(-a) = -f(a)$ for all $a \in R$.

We now prove some properties of ring homomorphisms.

Proposition 4.62. *Let $f \colon R \to R'$ be a ring homomorphism. Then:*

1. *If S is a subring of R, then $f(S)$ is a subring of R'.*

2. *If S' is a subring of R', then $f^{-1}(S')$ is a subring of R.*

3. *If I is an ideal of R, then $f(I)$ is an ideal of $f(R)$. [Not of R'!]*

4. *If I' is an ideal of R', then $f^{-1}(I')$ is an ideal of R.*

5. *If $g\colon R' \to R''$ is a ring homomorphism, then $g \circ f\colon R \to R''$ is a ring homomorphism.*

6. *Assume $f(1_R) = 1_{R'}$. If $a \in R$ is an invertible element, then $f(a^{-1}) = f(a)^{-1}$.*

Proof. The proofs of (1)-(5) are similar to the corresponding properties in Proposition 1.71 for groups. We have to use the subring or ideal criterion and the definition of ring homomorphism. For what concerns (6), let $a \in R$ be multiplicately invertible; then

$$f(a)f(a^{-1}) = f(aa^{-1}) = f(1_R) = 1_{R'}.$$

Similarly $f(a^{-1})f(a) = 1_{R'}$. Hence $f(a)^{-1} = f(a^{-1})$. □

Remark 4.63. We showed in item (3) above that $f(I)$ is an ideal of $f(R)$ for any ideal I. In general $f(I)$ is not an ideal of R'. Consider for instance the ring homomorphism $\mathbb{Z} \hookrightarrow \mathbb{Q}$. Then $\mathbb{Z}$ is an ideal of $\mathbb{Z}$, but its image is not an ideal of $\mathbb{Q}$.

Definition 4.64. A bijective ring homomorphism is called a ***(ring) isomorphism*** and if there exists an isomorphism from R to R', then R and R' are ***isomorphic*** (denoted $R \cong R'$). An isomorphism from a ring to itself is called a ***(ring) automorphism***.

Remark 4.65. Just as for groups, the property of being isomorphic is an equivalence relation on the set of rings. This follows from the observations that:

- the identity map $R \to R$ is a ring isomorphism,

- the inverse of a ring isomorphism is a ring isomorphism,

- the composition of two ring isomorphisms is a ring isomorphism.

Example 4.66. The function $f\colon \mathbb{C} \to \mathbb{C}$ given by $f(x+iy) := x - iy$ is an automorphism.

Remark 4.67. Let $f\colon R \to R'$ be a ring homomorphism. The kernel of f, defined as

$$\ker(f) := \{a \in R \mid f(a) = 0_{R'}\} = f^{-1}(0_{R'}),$$

is an ideal of R by Proposition 4.62, part (4), since $\{0_{R'}\}$ is an *ideal* of R'. The image of f, that is,

$$\operatorname{im}(f) := \{f(a) \mid a \in R\} = f(R),$$

is a *subring* of R' by Proposition 4.62, part (1).

If f is regarded as a group homomorphism from $(R,+)$ to $(R',+)$, then one gets the same $\ker(f)$ and $\operatorname{im}(f)$. Hence the ring homomorphism $f\colon R \to R'$ is injective if and only if $\ker(f) = \{0_R\}$, by Proposition 1.80.

4.5 Quotients of rings

Recall that in Chapter 2 we studied quotients groups: given a group G and a normal subgroup $H \subseteq G$, the quotient group G/H is the set of cosets of H, and we defined a group structure on G/H. In this section we define the analogue notion in ring theory, that is, quotients rings.

In school you probably saw that to a polynomial equation such as $x^2+y^2-1=0$ one may associate a geometric object, that is, the set of points (x,y) in the plane that satisfy that equation. Algebraic geometry is a vast generalization of this. Quotient rings play an important role in algebraic geometry, where coordinate rings of *varieties* are quotient rings. A variety is a geometric object, whose geometric properties correspond to algebraic properties of the quotient ring associated to it.

Remarks 4.68.

- Let R be a ring, and let $I \subseteq R$ be an ideal. In particular I is a subgroup of $(R,+)$, and since $(R,+)$ is an abelian group (by the definition of a ring), I is automatically a normal subgroup. We may therefore consider the quotient group

$$R/I := \{a+I \mid a \in R\}$$

as in Chapter 2, that is, the set of cosets of I. By Theorem 2.43, R/I is an abelian group, with sum defined as

$$\begin{aligned} +\colon R/I \times R/I &\longrightarrow R/I \\ \big(a+I, b+I\big) &\longmapsto (a+b)+I. \end{aligned}$$

- Recall from the second item in Remarks 2.23, adjusted to this context, that the following equivalences hold:

$$x+I=y+I \quad\Leftrightarrow\quad x-y\in I \quad\Leftrightarrow\quad x\in y+I \quad\Leftrightarrow\quad y\in x+I.$$

Theorem 4.69. *Let R be a ring and I an ideal of R. Define the multiplication*

$$\begin{aligned} \cdot\colon R/I \times R/I &\longrightarrow R/I \\ (a+I, b+I) &\longmapsto ab+I. \end{aligned}$$

Then R/I, equipped with the sum and multiplication defined above, is a ring. Furthermore, the projection map

$$\begin{aligned} p_I\colon R &\longrightarrow R/I \\ a &\longmapsto a+I \end{aligned}$$

is a ring homomorphism.

Proof. First of all we show that the multiplication defined above is well-defined. Let $a+I = a'+I$ and $b+I = b'+I$. Then $a = a'+i_1$ and $b = b'+i_2$ for $i_1, i_2 \in I$, by the previous remark. We may then write

$$ab = (a'+i_1)(b'+i_2) = a'b' + a'i_2 + i_1b' + i_1i_2.$$

As I is an ideal, the elements $a'i_2$, i_1b' and i_1i_2 are all in I, and hence $a'i_2 + i_1b' + i_1i_2 \in I$. Therefore $ab+I = a'b'+I$, which means that the multiplication is well-defined.

Since sum and multiplication in R/I are defined using representatives from R, associativity of multiplication and distributivity follow from the corresponding properties in R.

Lastly, by Theorem 2.43, we already know that the map p_I is a group homomorphism. Moreover, we have

$$p_I(ab) = (ab) + I = (a+I)\cdot(b+I) = p_I(a)\cdot p_I(b),$$

where for the second equality we use the definition of multiplication in R/I. Hence p_I is a ring homomorphism. □

Definition 4.70. The ring R/I defined above is called a ***quotient ring***, or a ***factor ring***, of R by I, or modulo I.

Remarks 4.71.

- The additive identity of R/I is $0_R + I$, where 0_R is the additive identity of R. The additive inverse of $a+I$ is $-a+I$, and if a is a unit in R, then $a+I$ is a unit in R/I and its multiplicative inverse is $a^{-1}+I$.
- If R is a unital ring with unity 1_R, then R/I has unity $1_R + I$. In this case, note that p_I maps 1_R to $1_R + I$.
- If R is a commutative ring, then R/I is a commutative ring.
- An ideal in ring theory is analogous to a normal subgroup in the group theory as they are exactly the substructures that allow defining a quotient.

Example 4.72. Let $R = \mathbb{Z}$ and $I = n\mathbb{Z}$. The quotient ring is

$$\mathbb{Z}/n\mathbb{Z} = \{a + n\mathbb{Z} \mid a \in \mathbb{Z}\} = \{a + n\mathbb{Z} \mid a = 0, 1, \ldots, n-1\}$$

with the addition and multiplication defined as

$$(a+n\mathbb{Z}) + (b+n\mathbb{Z}) = (a+b) + n\mathbb{Z} \qquad \text{and} \qquad (a+n\mathbb{Z})\cdot(b+n\mathbb{Z}) = (ab) + n\mathbb{Z}.$$

This ring is isomorphic to $\mathbb{Z}_n$.

Remark 4.73. One thing to keep in mind in the following is that $x+I$ is the zero element of R/I if and only if $x \in I$ in R.

Proposition 4.74. *Let R be a commutative unital ring, and let $I \subset R$ be an ideal.*

(1) I is a prime ideal if and only if R/I is an integral domain,

(2) I is a maximal ideal if and only if R/I is a field.

Proof. In both cases, the fact that R/I is a nonzero ring (which is a requirement in the definitions of an integral domain and a field) is equivalent to the fact that I is a proper ideal (which is a requirement in the definitions of a prime and a maximal ideal), hence $1+I \neq 0+I$. Moreover, the commutativity of R/I (which is also a requirement in the definitions of an integral domain and a field) follows from that of R.

(1) Assume that I is a prime ideal. Let $a+I, b+I \in R/I$ be such that $(a+I)(b+I) = 0+I$, which means that $ab+I = I$, or in other words that $ab \in I$. Since I is a prime ideal, this implies that $a \in I$, in which case $a+I = 0+I$, or that $b \in I$, in which case $b+I = 0+I$. This shows that R/I is an integral domain.

Conversely, assume that R/I is an integral domain. Let $a, b \in R$ be such that $ab \in I$. This means that $ab+I = I = 0+I$, which we may rewrite as $(a+I)(b+I) = 0+I$. Since R/I is an integral domain, this implies that $a+I = 0+I$, in which case $a \in I$, or that $b+I = 0+I$, in which case $b \in I$. This shows that I is a prime ideal.

(2) Assume that I is a maximal ideal. Let $a+I$ in R/I be a nonzero element in R/I, which means that $a \in R \backslash I$. Since I is a maximal ideal and $a \notin I$, we necessarily have $R = (a)+I$. In particular, we may write $1 = ra+i$, for some $r \in R$ and some $i \in I$. When taking cosets, this means that

$$1 + I = (ra + i) + I = ra + I = (r + I)(a + I),$$

where in the second equality we used the fact that $i \in I$. But then $a+I$ has a multiplicative inverse, which is $r + I$, and thus R/I is a field.

Conversely, assume that R/I is a field. Let $J \supsetneq I$ be an ideal of I. In order to show that I is maximal, we will show that $J = R$. Since $J \supsetneq I$, there exists some $a \in J \setminus I$, hence in particular $a + I \neq 0 + I$ in R/I. Since R/I is a field, this guarantees that $a + I$ has a multiplicative inverse $b + I$, so that

$$ab + I = (a + I)(b + I) = 1 + I.$$

This in turn implies that $1 = ab + x$, for some $x \in I$. But then

$$1 = ab + x \in J + I \subseteq J + J = J,$$

and therefore $J = R$, which is what we needed to show.

□

Example 4.75. Consider $\mathbb{Z}$. As we saw in Example 4.57, the prime ideals of $\mathbb{Z}$ are (0) and (p), where p is a prime number. The ideals of the form (p) for a prime number p are in fact maximal ideals. This tells us that the quotients of $\mathbb{Z}$ that are integral domains are exactly $\mathbb{Z}/(0) \cong \mathbb{Z}$ and the $\mathbb{Z}_p$'s, and the quotients of $\mathbb{Z}$ that are fields are exactly the $\mathbb{Z}_p$'s.

For instance, 6 is not a prime number, and indeed $\mathbb{Z}_6$ is not an integral domain, as $\overline{2} \cdot \overline{3} = \overline{0}$.

4.5.1 First isomorphism theorem for rings

The isomorphism theorems for rings are similar to the ones for groups with the notion of a normal subgroup replaced by the notion of an ideal. We will state and prove only the First Isomorphism Theorem, which is by far the most important.

Theorem 4.76 (First Isomorphism Theorem for Rings). *Let $f\colon R \to R'$ be a ring homomorphism. Then*

$$R/\ker(f) \cong \operatorname{im}(f).$$

Proof. By the First Isomorphism Theorem for Groups (Theorem 2.49), the map

$$\begin{aligned} \varphi\colon R/\ker(f) &\longrightarrow \operatorname{im}(f) \\ a + \ker(f) &\longmapsto f(a) \end{aligned}$$

is a group isomorphism. It is also a ring isomorphism, because

$$\varphi((a+\ker(f))(b+\ker(f))) = \varphi((ab)+\ker(f)) = f(ab) = f(a)f(b) = \varphi(a+\ker(f))\varphi(b+\ker(f)).$$

□

In summary, by Theorem 4.69, every quotient ring R/I gives rise to a ring homomorphism p_I, and by Theorem 4.76, every ring homomorphism f gives rise to a quotient ring $R/\ker(f)$. The images of ring homomorphisms from R are isomorphic to factor rings of R.

Examples 4.77. 1. The map

$$\begin{aligned} f\colon \mathbb{Z} &\longrightarrow \mathbb{Z}_n \\ a &\longmapsto [a]_n, \end{aligned}$$

where $[a]_n$ is the remainder of a on division by n, is a surjective ring homomorphism. The kernel of f is $\ker(f) = n\mathbb{Z}$. By the first isomorphism theorem,

$$\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n.$$

2. The map

$$\begin{aligned} f\colon \mathbb{R}[x] &\to \mathbb{R} \\ a_0 + a_1x + \cdots + a_nx^n &\longmapsto a_0 \end{aligned}$$

is a surjective ring homomorphism. (It can be alternatively described as the evaluation of a polynomial at $x = 0$, which would then make it a special case of the second item in Examples 4.60.) The kernel of f is

$$\ker(f) = (x) = \{a_1x + \cdots + a_nx^n \mid n \geq 1, a_i \in \mathbb{R} \text{ for all } i\}.$$

By the first isomorphism theorem, $\mathbb{R}[x]/(x) \cong \mathbb{R}$.

4.5.2 The correspondence theorem

Given a ring R and two ideals I and J with $I \subseteq J \subseteq R$, we denote

$$J/I := \{a + I \mid a \in J\}.$$

Theorem 4.78 (Correspondence Theorem). *Let R be a ring, and let $I \subseteq R$ be an ideal. The function*

$$\begin{aligned} \Phi\colon \{\textit{ideals of } R \textit{ containing } I\} &\longrightarrow \{\textit{ideals of } R/I\} \\ J &\longmapsto J/I \end{aligned}$$

is a bijection.

Proof. First of all we need to show that the map is well-defined, that is, that J/I is an ideal of R/I. We already know that J/I is a subgroup of R/I, by Theorem 2.59. What we are left to check is that, for all $a + I \in J/I$ and all $r + I \in R/I$, the products $(a + I)(r + I)$ and $(r + I)(a + I)$ are in J/I. Indeed, since J is an ideal, we know that $ar \in J$ and $ra \in J$, so that

$$(a + I)(r + I) = ar + I \in J/I \qquad \text{and} \qquad (r + I)(a + I) = ra + I \in J/I.$$

Next, we need to show injectivity and surjectivity of Φ. Exercise. □

Theorem 4.79. *Let R be a commutative ring, and let $I \subseteq R$ be an ideal. The function*

$$\begin{aligned} \Phi\colon \{\textit{prime ideals of } R \textit{ containing } I\} &\longrightarrow \{\textit{prime ideals of } R/I\} \\ P &\longmapsto P/I \end{aligned}$$

is a bijection.

4.5.3 Quotients of polynomial rings

We generalize the second item of Examples 4.77, which states that $\mathbb{R}[x]/(x) \cong \mathbb{R}$. First of all, we aim at a definition of a more general polynomial ring. Given a commutative ring R, the ***polynomial ring over*** R is

$$R[x] := \{a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + a_n x^n \mid n \in \mathbb{N},\ a_i \in R\}.$$

One may legitimately ask: what is x? Below we define formally what a polynomial is.

A polynomial is a sequence in R that is *eventually zero*, that is, a function $p\colon \mathbb{N} \to R$ such that there exists some $N \in \mathbb{N}$ for which, for all $i \geq N$, we have $p_(i) = 0_R$. A bit more explicitly, the polynomial p is a sequence

$$p = (p_0, p_1, p_2, \dots),$$

where $p_i = 0_R$ for all $i \geq N$. For two such sequences p and q, one defines

$$p + q := (p_0 + q_0, p_1 + q_1, p_2 + q_2, \dots),$$

or more formally, $p + q\colon \mathbb{N} \to R$ is defined by

$$(p+q)(i) := p(i) + q(i).$$

Since p and q are both eventually zero, also $p+q$ is eventually zero. The neutral element with respect to $+$, that is, the zero of this ring, is $(0,0,0,\dots)$. One may also define a product $p \cdot q$, which in the "semi-formal way" is written

$$p \cdot q := (p_0 q_0, p_1 q_0 + p_0 q_1, p_2 q_0 + p_1 q_1 + p_0 q_2, \dots),$$

and more formally is given by

$$(p \cdot q)(i) := \sum_{k=0}^{i} p(i-k) q(k).$$

The unity of this ring is $(1,0,0,\dots)$. For instance, if $R = \mathbb{R}$, we may perform the following operations on polynomials:

$$\begin{aligned}(2,1,3,0,0,\dots) + (1,4,0,0,\dots) &= (3,5,3,0,0,\dots),\\ (2,1,3,0,0,\dots) \cdot (1,4,0,0,\dots) &= (2,\ 1+8,\ 3+4+0,\ 0+12+0+0,\ 0,0,\dots)\\ &= (2,9,7,12,0,0,\dots).\end{aligned}$$

These might look rather confusing, so here comes the brilliant idea: define

$$x := (0,1,0,0,0,\dots).$$

Then, you may check that

$$\begin{aligned}x \cdot x &= (0,0,1,0,0,0,\dots),\\ x \cdot x \cdot x &= (0,0,0,1,0,0,0,\dots),\\ x \cdot x \cdot x \cdot x &= (0,0,0,0,1,0,0,0,\dots),\end{aligned}$$

and so on. One usually writes these polynomials as x^2, x^3 and x^4. Moreover, if we agree that x^0 should be the unity $(1,0,0,\dots)$ of this polynomial ring, and we formally write how to multiply a polynomial by a constant, one may easily "decompose" for instance

$$\begin{aligned}(2,5,3,0,0,\dots) &= (2,0,0,\dots) + (0,5,0,0,\dots) + (0,0,3,0,0,\dots)\\ &= 2(1,0,0,\dots) + 5(0,1,0,0,\dots) + 3(0,0,1,0,0,\dots)\\ &= 2x^0 + 5x + 3x^2,\end{aligned}$$

and this is a much more familiar way of writing polynomials. End of the digression.

As stated in the beginning of this section, we now aim for some isomorphism theorems for polynomial rings. First, an important fact useful in what follows:

Theorem 4.80. *Let R be a commutative unital ring, and consider the polynomial ring $R[x]$.*

1. *(The Euclidean division with remainder for polynomials.) Let f and g be polynomials in $R[x]$ with $f = a_0 + a_1x + \cdots + a_nx^n$, and assume that the coefficient a_n is an invertible element of R. There there are unique polynomials q and p in $R[x]$ that satisfy*

$$g = q \cdot f + p, \qquad \text{with } p = 0 \text{ or } \deg(p) < \deg(f).$$

2. *As a consequence, each ideal of $K[x]$ is principal, where K is a field.*

(The reason for the "$p = 0$ or..." in part 1 is that often people prefer not to define the degree of the zero polynomial. This depends a lot on the authors and the context.)

Example 4.81. For instance $g := x^2 + 1$ cannot be divided by $f := 2x + 1$ in $\mathbb{Z}[x]$, since 2 is not invertible in $\mathbb{Z}$.

Lemma 4.82. *Let $\varphi\colon A \to B$ be a surjective ring homomorphism, and let $I \subseteq A$ be an ideal. Then*

$$A/I \cong B/\varphi(I).$$

Proof. First of all, $\varphi(I)$ is an ideal of B by item 3 of Proposition 4.62. Consider now the composition

$$A \xrightarrow{\varphi} B \xrightarrow{p} B/\varphi(I),$$

where p is the projection. Then $p \circ \varphi$ is a surjective ring homomorphism, and $\ker(\pi \circ \varphi) = I$. We conclude by the first isomorphism theorem. □

Theorem 4.83. *Let R be a commutative unital ring. Let $a \in R$.*

1. *Consider the principal ideal $(x - a)$ in $R[x]$. We have*

$$R[x]/(x - a) \cong R.$$

2. *Let $f_1, f_2, \dots, f_s \in R[x]$ be polynomials, and consider the ideal $I := (x - a, f_1, \dots, f_s)$ in $R[x]$. Then*

$$R[x]/I \cong R/\big(f_1(a), \dots, f_s(a)\big).$$

Proof. (1) Consider the map

$$\begin{aligned} \varphi\colon R[x] &\longrightarrow R \\ f &\longmapsto f(a). \end{aligned}$$

This is a surjective ring homomorphism. We show that $\ker(\varphi) = (x - a)$:

($\supseteq$) The elements of the principal ideal $(x - a)$ are the multiples of $x - a$, that is, the polynomials of the form $(x - a) \cdot f$, for $f \in R[x]$. Any such multiple evaluated at a vanishes, hence it is in $\ker(\varphi)$.

($\subseteq$) Let $g \in \ker(\varphi)$. Then $g(a) = 0$. By Theorem 4.80, we may write

$$g = (x-a)\cdot q + p,$$

with $p = 0$ or $\deg(p) < 1$, which means that p is a constant. If we now evaluate both sides at a, we get

$$g(a) = (a-a)\cdot q(a) + p,$$

which simplifies to $0 = p$, so that g is actually a multiple of $x-a$.

(2) *Sketch.* One may prove an additional isomorphism theorem for rings, similar to the third isomorphism theorem for groups, stating that if we have ideals J_1 and J_2 in a ring A with $J_1 \subseteq J_2 \subseteq A$, then

$$(A/J_1)\big/(J_2/J_1) \cong A/J_2.$$

We may apply this to the present theorem, since we have $(x-a) \subseteq I$, and get

$$\big(R[x]/(x-a)\big)\big/\big(I/(x-a)\big) \cong R[x]/I.$$

On the other hand, we know that

$$\big(R[x]/(x-a)\big)\big/\big(I/(x-a)\big) \cong R\big/\big(I/(x-a)\big) \cong R/\big(f_1(a),\dots,f_s(a)\big),$$

where for the first isomorphism we use part (1), and the second isomorphism follows from Lemma 4.82. □

Examples 4.84. 1. To illustrate part (1) of the theorem above, for instance we have

$$\mathbb{Z}[x]/(x-3) \cong \mathbb{Z}, \qquad \mathbb{Z}[x]/(x-5) \cong \mathbb{Z}.$$

We can even use it in wilder ways. For instance with a polynomial ring in *two* variables:

$$\mathbb{Q}[x,y]/(x+y^2) \cong \mathbb{Q}[y].$$

Indeed it is true that $\mathbb{Q}[x][y] \cong \mathbb{Q}[y][x] \cong \mathbb{Q}[x,y]$.

2. To illustrate (2),

$$\mathbb{Z}[x]\big/(x-3,\, x^2+1) \cong \mathbb{Z}/(3^2+1) \cong \mathbb{Z}_{10}.$$

We conclude this section with a remark on the importance of polynomial rings. In group theory, Corollary 3.35 shows that every finitely generated abelian group is isomorphic to a quotient of $\mathbb{Z}^n$, for a suitable n. Therefore, in order to understand all finitely generated abelian groups, however abstract they might seem, it is actually enough to understand $\mathbb{Z}^n$ and its quotients. Polynomial rings play a similar role in ring theory, as stated in Corollary 4.86 below. The results are best stated in terms of K-algebras: essentially, given a field K, a K-***algebra*** is a commutative ring R that contains K as a subring. (There are more sophisticated definitions.) For instance, a polynomial ring in any number of variables $K[x_1,x_2,\dots,x_n]$ is a K-algebra: it is a commutative ring that contains K as a subring, namely the constant polynomials.

Theorem 4.85 (Universal property of polynomial rings)**.** *Let R be a K algebra, and let $y_1,\dots,y_n \in R$. There is a unique ring homomorphism $\varphi\colon K[x_1,\dots,x_n]$ with $\varphi(x_i) = y_i$ for all $i \in \{1,\dots,n\}$ and such that $\varphi(a) = a$ for all $a \in K$.*

Proof. The proof is similar to that of Proposition 3.33. □

We say that a K-algebra R is ***finitely generated*** if there is a finite set $\{y_1,\dots,y_n\}$ of elements of R such that all the elements of R can be written as a polynomial combinations of the y_i's, with coefficients in K.

Corollary 4.86. *Every finitely generated K-algebra is isomorphic to a quotient of $K[x_1,\dots,x_n]$, for a suitable n.*

Proof. If R is a finitely generated K-algebra and $y_1,\dots,y_n$ are generators for R, we may consider the map $\varphi\colon K[x_1,\dots,x_n] \to R$ of Theorem 4.85. This is a surjective ring homomorphism, and therefore

$$R \cong K[x_1,\dots,x_n]/\ker(\varphi)$$

by the first isomorphism theorem. □

4.6 Field extensions

The following example is *not* a special case of Theorem 4.83. It is worth remembering this one example, even if you want to forget everything else about quotients of rings.

Example 4.87 (Important!)**.** Consider the polynomial ring $\mathbb{R}[x]$ over the reals. Then

$$\mathbb{R}[x]/(x^2+1) \cong \mathbb{C}.$$

Proof. Intuitively, the ideal we quotient by is what becomes zero in the quotient. So inside $\mathbb{R}[x]/(x^2+1)$ we have $x^2+1=0$, which means that $x^2=-1$. This definitely looks a lot like what happens with complex numbers, where $i^2=-1$. More formally, consider the function

$$\begin{aligned}\psi\colon \mathbb{R}[x] &\longrightarrow \mathbb{C}\\ f &\longmapsto f(i).\end{aligned}$$

This is a ring homomorphism. Moreover, we show that $\ker(\psi)=(x^2+1)$:

($\supseteq$) The elements of the principal ideal (x^2+1) are the multiples of x^2+1, that is, of the form $(x^2+1)\cdot f$, for $f\in\mathbb{R}[x]$. Of course, any such multiple evaluated at i vanishes, hence it is in $\ker(\psi)$.

($\subseteq$) Let $g\in\ker(\psi)$. Then $g(i)=0$. Write

$$g=(x^2+1)\cdot q+p,$$

with $p=0$ or $\deg(p)<2$, which means that p is a polynomial of degree 1 or a constant. Since we know that $g(i)=0$, the right-hand side evaluated at i also has to vanish, which means that $p(i)=0$. But there is no polynomial in $\mathbb{R}[x]$ of degree 1 that vanishes at i, hence necessarily $p=0$. Therefore $g=(x^2+1)\cdot q$, so that $g\in(x^2+1)$.

Lastly, the map ψ is surjective. Hence, the statement follows from the first isomorphism theorem. □

In the rest of this section, our goal is to generalize this example.

Definition 4.88. Let K and L be fields, with $K\subseteq L$. We call such an inclusion of fields a ***field extension***. Let $\alpha\in L$. Define

$$\begin{aligned}\psi_\alpha\colon K[x] &\longrightarrow L\\ f &\longmapsto f(\alpha),\end{aligned} \qquad\qquad K[\alpha]:=\mathrm{im}(\psi_\alpha).$$

It is easy to see that the map ψ_α is a ring homomorphism.

Definition 4.89. Let K and L be fields, with $K \subseteq L$. Let $\alpha \in L$.

- We say that α is ***algebraic over*** K if $\ker(\psi_\alpha) \neq (0)$, that is, if there exists some nonzero polynomial $f \in K[x]$ such that $f(\alpha) = 0$.

- We say that α is ***transcendental over*** K if $\ker(\psi_\alpha) = (0)$, that is, if the only polynomial $f \in K[x]$ such that $f(\alpha) = 0$ is $f = 0$.

Examples 4.90. 1. Consider $\mathbb{R} \subset \mathbb{C}$. The imaginary number i is algebraic over $\mathbb{R}$, since the polynomial $f := x^2 + 1 \in \mathbb{R}[x]$ is such that $f(i) = 0$. We have $\mathbb{R}[i] = \mathbb{C}$.

2. Consider $\mathbb{Q} \subset \mathbb{R}$. The irrational number $\sqrt{2}$ is algebraic over $\mathbb{Q}$, since the polynomial $f := x^2 - 2 \in \mathbb{Q}[x]$ is such that $f(\sqrt{2}) = 0$.

3. Some famous results by Lindemann and Hermite state that the real numbers π and e are transcendental over $\mathbb{Q}$: there is no polynomial in $\mathbb{Q}[x]$ that vanishes at π or e.

Remark 4.91. By Theorem 4.80, since $\ker(\psi_\alpha)$ is an ideal of $K[x]$, it is in fact a principal ideal, so that $\ker(\psi_\alpha) = (f)$ for some $f \in K[x]$. The generator f is not uniquely determined, because we may always multiply it by constants in K without changing the ideal. But the generator *is* uniquely determined if we assume that it is ***monic***, that is, that the coefficient of the term of highest degree is equal to $1 = 1_K$. This unique generator is the monic polynomial of smallest degree vanishing at α.

Definition 4.92. With the same notation as above, assuming that α is algebraic over K, the unique monic generator of $\ker(\alpha)$ is called the ***minimal polynomial of α over K***.

Examples 4.93. 1. The minimal polynomial of i over $\mathbb{R}$ is $x^2 + 1$.

2. The minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$.

3. The minimal polynomial of $\sqrt{3}$ over $\mathbb{Q}$ is $x^2 - 3$.

Proposition 4.94. *Let $K \subseteq L$ be a field extension. If $\alpha \in L$ is algebraic over K, then $K[\alpha]$ is a field.*

Proof. By the first isomorphism theorem, we get that

$$K[x]/\ker(\psi_\alpha) \cong K[\alpha],$$

and since $K[\alpha]$ is a subring of L, which is a field and hence an integral domain, $K[\alpha]$ is itself an integral domain. By Proposition 4.74, this is equivalent to $\ker(\psi_\alpha)$ being a prime ideal of $K[x]$. The minimal polynomial f of α is therefore an irreducible polynomial. But then $\ker(\psi_\alpha) = (f)$ is in fact a maximal ideal: if g is a polynomial $K[x] \setminus (f)$, then $(f, g) = (1) = K[x]$. This can be seen analogously to an analogous argument to the one mentioned in Examples 4.44 for $\mathbb{Z}$, by using the division with remainder in $K[x]$. To conclude, again by Proposition 4.74, since $\ker(\psi_\alpha) = (f)$ is a maximal ideal, $K[\alpha]$ is a field. □

Remark 4.95. • In fact, $K[\alpha]$ is a field if and only if α is algebraic: if instead α is transcendental, then $\ker(\psi_\alpha) = (0)$, and

$$K[\alpha] \cong K[x]/(0) \cong K[x]$$

is an integral domain but not a field, where the first isomorphism comes from the first isomorphism theorem.

- The previous proposition gives us a way to construct new fields starting from a given field K, by "adjoining algebraic elements" to K.

Examples 4.96. 1. Consider $\mathbb{R} \subset \mathbb{C}$ and $i \in \mathbb{C}$. As previously observed, $\mathbb{R}[i] = \mathbb{C}$. Let us elaborate on this. Simply by definition of image, we may write

$$\begin{aligned}\mathbb{R}[i] &= \{f(i) \mid f \in \mathbb{R}[x]\} \\ &= \{a_0 + a_1 i + a_2 i^2 + \cdots + a_n i^n \mid n \in \mathbb{N}, a_i \in \mathbb{R}\},\end{aligned}$$

but since all the powers i^k for $k \geq 2$ can be expressed in terms of the zeroth and first powers of i, we may simply rewrite

$$\mathbb{R}[i] = \{b_0 + b_1 i \mid b_0, b_1 \in \mathbb{R}\} = \mathbb{C}.$$

2. Consider $\mathbb{Q} \subset \mathbb{R}$ and $\sqrt{2} \in \mathbb{R}$. Similarly to the previous point, we may write

$$\begin{aligned}\mathbb{Q}[\sqrt{2}] &= \{f(\sqrt{2}) \mid f \in \mathbb{Q}[x]\} \\ &= \{a_0 + a_1\sqrt{2} + a_2(\sqrt{2})^2 + \cdots + a_n(\sqrt{2})^n \mid n \in \mathbb{N}, a_i \in \mathbb{Q}\} \\ &= \{b_0 + b_1\sqrt{2} \mid b_0, b_1 \in \mathbb{Q}\},\end{aligned}$$

where in the last equality we use the fact that all high powers of $\sqrt{2}$ can be expressed in terms of the zeroth and first powers of $\sqrt{2}$. By the proposition above, we know that $\mathbb{Q}[\sqrt{2}]$ is a field, so in particular every nonzero element has a multiplicative inverse. For instance, we claim that the inverse of $3 - 5\sqrt{2}$ is $\frac{3}{-41} + \frac{5}{-41}\sqrt{2}$: indeed

$$(3-5\sqrt{2})\left(\frac{3}{-41} + \frac{5}{-41}\sqrt{2}\right) = \frac{1}{-41}(3-5\sqrt{2})(3+5\sqrt{2}) = \frac{1}{-41}(3^2 - (5\sqrt{2})^2) = \frac{1}{-41}(-41) = 1.$$

(The middle-school result that $(a+b)(a-b) = a^2 - b^2$ is useful for finding inverses in this field.)

3. Consider $\mathbb{Q} \subset \mathbb{R}$ and $\sqrt[3]{2} \in \mathbb{R}$. Now we need the zeroth, first and second powers of $\sqrt[3]{2}$ in order to express all other powers:

$$\begin{aligned}\mathbb{Q}[\sqrt[3]{2}] &= \{f(\sqrt[3]{2}) \mid f \in \mathbb{Q}[x]\} \\ &= \{a_0 + a_1\sqrt[3]{2} + a_2(\sqrt[3]{2})^2 + \cdots + a_n(\sqrt[3]{2})^n \mid n \in \mathbb{N}, a_i \in \mathbb{Q}\} \\ &= \{b_0 + b_1\sqrt[3]{2} + b_2(\sqrt[3]{2})^2 \mid b_0, b_1, b_2 \in \mathbb{Q}\}.\end{aligned}$$

4. As you might guess, if we consider some n-th root then we need n terms. For instance,

$$\mathbb{Q}[\sqrt[n]{2}] = \{b_0 + b_1\sqrt[n]{2} + b_2(\sqrt[n]{2})^2 + \cdots + b_{n-1}(\sqrt[n]{2})^{n-1} \mid b_0, b_1, b_2, \ldots, b_{n-1} \in \mathbb{Q}\}.$$

5. Consider $\mathbb{Q} \subset \mathbb{R}$ and $\pi \in \mathbb{R}$. There is no polynomial with rational coefficients vanishing at π, and hence $\mathbb{Q}[\pi]$ is not a field. Unlike the previous examples, there is no expression for $\mathbb{Q}[\pi]$ simpler than

$$\mathbb{Q}[\pi] = \{a_0 + a_1\pi + a_2\pi^2 + \cdots + a_n\pi^n \mid n \in \mathbb{N}, a_i \in \mathbb{Q}\}.$$

6. Consider a field K and $\alpha \in K$. There is a monic linear polynomial in $K[x]$ that vanishes at α, namely $x-\alpha$. This is the minimal polynomial of α over K, that is, $\ker(\psi_\alpha) = (x-\alpha)$. In this case, we get

$$K[\alpha] \cong K[x]/(x-\alpha) \cong K,$$

where the second isomorphism follows from Theorem 4.83.

We do not have the necessary terminology and tools to properly discuss and prove some of the following statements, so we simply give them without proof.

Proposition 4.97. *Consider a field extension $K \subseteq L$. Let $\alpha, \beta \in L$ be algebraic over K. Then*

$$K[\alpha][\beta] \cong K[\beta][\alpha].$$

We simply denote this field as $K[\alpha, \beta]$.

Example 4.98. Consider $\mathbb{Q} \subset \mathbb{R}$ and the elements $\sqrt{2}$ and $\sqrt{3}$ in $\mathbb{R}$. One may prove, by contradiction, that $\sqrt{3} \notin \mathbb{Q}[\sqrt{2}]$, so that there are strict inclusions

$$\mathbb{Q} \subsetneq \mathbb{Q}[\sqrt{2}] \subsetneq \mathbb{Q}[\sqrt{2}, \sqrt{3}].$$

Theorem 4.99. *Consider a field extension $K \subseteq L$. Let $\alpha, \beta \in L$ be algebraic elements over K. Then $\alpha + \beta$ and $\alpha\beta$ are also algebraic elements. If $\alpha \neq 0$, then α^{-1} is an algebraic element. In particular, the subset of L*

$$\overline{K}^L := \{\alpha \in L \mid \alpha \text{ is algebraic over } K\}$$

*is a field, called the **algebraic closure of** K **in** L.*

Examples 4.100. 1. We have strict inclusions $\mathbb{Q} \subsetneq \overline{\mathbb{Q}}^{\mathbb{R}} \subsetneq \mathbb{R}$.

2. Consider $\mathbb{Q} \subset \mathbb{R}$ and $\mathbb{Q} \subsetneq \mathbb{Q}[\sqrt{2}] \subsetneq \mathbb{Q}[\sqrt{2}, \sqrt{3}]$. By the theorem, since both $\sqrt{2}$ and $\sqrt{3}$ are algebraic over $\mathbb{Q}$, also their sum $\sqrt{2} + \sqrt{3}$ is algebraic over $\mathbb{Q}$. Let us find the minimal polynomial of $\sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$. One may observe that

$$(\sqrt{2} + \sqrt{3})^2 = 5 + 2\sqrt{6} \qquad \text{and} \qquad \left((\sqrt{2} + \sqrt{3})^2 - 5\right)^2 = 24,$$

and by suitable combining these one finds that the polynomial

$$(x^2 - 5)^2 - 24$$

has $\sqrt{2} + \sqrt{3}$ as a root. By using some appropriate result on the irreducibility of polynomials with rational coefficients, one may check that this polynomial is irreducible in $\mathbb{Q}[x]$, and it is therefore the minimal polynomial of $\sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$.

Definition 4.101. Let K be a field.

- A field extension $K \subseteq L$ is called an ***algebraic extension*** if $\overline{K}^L = L$.
- We call K an ***algebraically closed field*** if K does not have proper algebraic extensions.
- An ***algebraic closure of*** K is an algebraic extension L of K that is algebraically closed.

Theorem 4.102. *Every field K has an algebraic closure, which is unique up to isomorphism and denoted $\overline{K}$.*

Examples 4.103. We have $\overline{\mathbb{R}} = \mathbb{C}$ and $\overline{\mathbb{Q}} \subsetneq \mathbb{C}$.

Chapter 5

What next?

This chapter does not introduce anything new needed for this course. It is only meant to provide some pointers to other courses and branches of math where the topics introduced in Abstract Algebra are used. Some courses offered at Aalto that make direct use of groups and/or rings are:

- Galois theory (current year, period 4),
- Commutative algebra (current year, period 4),
- Algebraic number theory (current year, period 5),
- Real algebraic geometry (current year, period 5),
- Computational algebraic geometry (every other year, next time in 2023-2024),
- Lie groups and Lie algebras (every other year, next time in 2023-2024).

Below follow some very brief introductions to some areas and topics, most of which are related to the courses listed above. A good point to keep in mind is that, historically,

the main goal in algebra is finding the solutions of polynomial equations.

Different areas deal with different kinds of equations. This may sound much simpler than the topics covered in this course, but the notions of group, ring, etc., where originally motivated by such problems.

Linear algebra, functional analysis, etc

The main topic of study in linear algebra is systems of linear equations. One may approach this from more or less abstract points of view.

In Abstract Algebra we dealt mostly with binary operations on a set, that is, operations that take two elements of that set as input and return an element of that same set. It is often natural to define a different kind of operation, namely a "multiplication by scalars":

$$\cdot\colon K \times V \longrightarrow V.$$

If K is a field and V is an abelian group, such a multiplication by scalars has to satisfy some conditions similar to those in the definition of a ring. For instance is $V = \mathbb{R}^n$ consists of vectors with n real entries under coordinate-wise addition, the multiplication by scalars in $K = \mathbb{R}$ is exactly the one you saw in matrix algebra. Or one could take the abelian group of matrices $V = M_{n,m}(\mathbb{R})$ with sum, and equip it with the usual multiplication by scalars in $K = \mathbb{R}$. These are examples of ***vector spaces***, the algebraic structure studied in linear algebra.

Depending on personal preferences and other courses in the curricula, linear algebra may be taught from a different perspective: less abstract algebra and more norms. This approach is closer to functional analysis than abstract algebra.

Algebraic geometry

In primary school we learn how to compute the area of a rectangle by multiplying the lengths of the sides. In high school we learn that equations can describe geometric objects, for instance the points (x,y) in the plane that satisfy $x^2 + y^2 = 1$ are the points on a circle of center the origin and radius 1. Algebraic geometry is the next step on this road. There are at least two kinds of algebraic geometry:

- *Classical algebraic geometry.* A rather intuitive generalization of the ideas learnt in high school: instead of considering geometric objects defined by just one equation (e.g., a circle), one considers systems of polynomial equations. This gives an ideal in a polynomial ring (because if two polynomials vanish, then their sum vanishes, too, and if a polynomial vanishes and you multiply it by any other polynomial, the result still vanishes). And geometric operations on the geometric object correspond to algebraic operations on the ideal.

- *Scheme theory.* A very abstract generalization developed in the second part of the last century. This machinery provides a unified approach to problems in classical algebraic geometry and number theory.

There are even more abstract constructions. Algebraic geometry is generally regarded as difficult, because it is connected to many areas of math.

References: *Basic Algebraic Geometry* by Shafarevich, *Ideals, Varieties and Algorithms* by Cox–Little–O'Shea, *Algebraic Geometry* by Hartshorne, *The Rising Sea* by Vakil.

Commutative algebra

Commutative algebra is the study of commutative rings. Given a commutative ring R, an R-***module*** is a triple $(M,+,\cdot)$, where $(M,+)$ is an abelian group and

$$\cdot\colon R \times M \longrightarrow M$$

is a "multiplication by scalars" that satisfies some conditions, similarly to the multiplication by scalars in a vector space. In fact, when R is a field, the notion of an R-module specializes exactly to that of a vector space. And on the opposite end, when one takes $R = \mathbb{Z}$, the corresponding notion of $\mathbb{Z}$-module is exactly the same as that of an abelian group.

The study of such structures is primarily motivated by the interplay with algebraic geometry, but they are also used in algebraic number theory and elsewhere.

References: *Commutative Algebra with a View towards Algebraic Geometry* by Eisenbud, *Introduction to Commutative Algebra* by Atiyah–Macdonald.

Computational algebra

Computational algebra deals with the actual problem of finding solutions to polynomial equation systems, or related problems, in an algorithmic way. For instance, one of the first and most basic problems is the following: given a polynomial $f \in \mathbb{Q}[x_1, \dots, x_n]$ and an ideal I in the same ring, how can one determine whether f belongs to I? Such a question is answered algorithmically by first finding what is called a ***Gröbner basis*** of I, which is a special kind of set of generators for I.

This is called Computational Algebraic Geometry in Aalto, since it is taught along with applications to algebraic geometry.

References: *Ideals, Varieties and Algorithms* by Cox–Little–O'Shea.

Algebraic number theory

As stated above, the goal in algebra is to solve polynomial equations. Perhaps surprisingly, the simpler the ring of coefficients is, the harder it is to find solutions. For instance, a univariate polynomial of degree n with coefficients in $\mathbb{C}$ has exactly n roots, counted with multiplicity. But there are polynomials with coefficients in $\mathbb{R}$ that do not have roots in $\mathbb{R}$, like $x^2 + 1$, or similarly $x^2 - 2$ over $\mathbb{Q}$, and so on. Polynomials with coefficients in $\mathbb{Z}$ are dealt with in number theory.

To be reductionist, number theory is the study of prime numbers. Analytic number theory approaches this topic with the use of analytic methods (complex analysis, etc.), whereas algebraic number theory uses the language of rings and ideals.

Algebraic topology

Topology is a branch of math that deals with geometric properties that are preserved under "continuous deformations" of a space.

Definition 5.1. A ***topological space*** is a pair (X, τ), where X is a set and τ is a family of subsets of X satisfying the following properties:

1. $\emptyset \in \tau$ and $X \in \tau$;
2. if I is any index set and for all $i \in I$ we have $U_i \in \tau$, then $\bigcup_{i \in I} U_i \in \tau$;
3. if $U \in \tau$ and $V \in \tau$, then $U \cap V \in \tau$.

With these assumptions we say that τ is a ***topology on*** X.

The pure study of topological spaces is classically called "point-set topology" or "general topology". There are constructions that associate groups to a topological space, and the study of such constructions is called "algebraic topology". It turns out that geometric properties of the topological space correspond to algebraic properties of the associated groups.

References: *A First Course in Algebraic Topology* by Kosniowski, *Singular Homology Theory* by Massey.