

II The Integers

1. Semigroups and Monoids

In Chapter I, we learned about the natural numbers with the operations of addition and multiplication. We may think about addition and multiplication as processes whereby we take two natural numbers m_1, m_2 and form another natural number, namely the sum $m_1 + m_2$ or the product $m_1 \cdot m_2$. We may formalize this idea by saying that the set of natural numbers has defined on it the *operations* $+$ and \cdot that assign to two natural numbers m_1, m_2 the respective natural numbers $m_1 + m_2$ and $m_1 \cdot m_2$. If we let $\mathbb{N} \times \mathbb{N}$ denote the set of all ordered pairs of natural numbers, called the *Cartesian product* of \mathbb{N} with itself, we may consider the operations of addition and multiplication to be mappings from $\mathbb{N} \times \mathbb{N}$ to \mathbb{N} given by the assignments $(m_1, m_2) \mapsto m_1 + m_2$ and $(m_1, m_2) \mapsto m_1 \cdot m_2$.

In what follows, we shall investigate the idea of a nonempty set M on which an operation \circ_M is defined. In this case, we have a mapping from $M \times M$ to M given by the assignment $(m_1, m_2) \mapsto m_1 \circ_M m_2$.

Generalizing the associativity of addition and multiplication of natural numbers, we call an operation \circ_M on a set M *associative* if for all elements m_1, m_2, m_3 of M , we have

$$(m_1 \circ_M m_2) \circ_M m_3 = m_1 \circ_M (m_2 \circ_M m_3).$$

If we are given an associative operation \circ_M on M , we can perform the operation on three elements m_1, m_2, m_3 either by operating on the first two and then operating on that result with the third, or by operating on the last two and then operating on the first with that result. We may therefore write simply $m_1 \circ_M m_2 \circ_M m_3$.

Definition 1.1. A nonempty set H with an associative operation \circ_H is called a *semigroup*.

For such a semigroup, we write (H, \circ_H) . If the context makes clear the connection with H , we may write simply (H, \circ) . If it is clear that we are dealing with a semigroup, we may suppress reference to the operation and write simply H .

Example 1.2. (i) The natural numbers \mathbb{N} with addition and with multiplication forms the semigroups $(\mathbb{N}, +)$ and (\mathbb{N}, \cdot) .

(ii) Let A be an arbitrary nonempty set. On the set

$$\text{map}(A) := \{f \mid f: A \longrightarrow A\}$$

of all mappings of A to itself, we define the operation \circ of composition of mappings, which is well known to be associative. With this operation, $(\text{map}(A), \circ)$ is a semigroup.

(iii) Let n be a nonzero natural number. Consider the subset

$$\mathcal{R}_n := \{0, \dots, n-1\}$$

of the natural numbers comprising the first n natural numbers. On the set \mathcal{R}_n , we can define two operations. Let us denote by $R_n(c)$ the remainder resulting from the division of a natural number c by n ; that this number is uniquely determined is guaranteed by Theorem 5.1 of Chapter I. We observe that we have $R_n(c) \in \mathcal{R}_n$. For two numbers $a, b \in \mathcal{R}_n$, we now define the mappings

$$\oplus : \mathcal{R}_n \times \mathcal{R}_n \longrightarrow \mathcal{R}_n, \text{ defined by } a \oplus b := R_n(a + b), \quad (1)$$

$$\odot : \mathcal{R}_n \times \mathcal{R}_n \longrightarrow \mathcal{R}_n, \text{ defined by } a \odot b := R_n(a \cdot b). \quad (2)$$

We leave it as an exercise for the reader to verify that the operations \oplus and \odot are associative (the associativity is derived from the associativity of addition and multiplication on the set of natural numbers). We thereby obtain two semigroups, (\mathcal{R}_n, \oplus) and (\mathcal{R}_n, \odot) .

Exercise 1.3. Verify that the operations \oplus and \odot from Example 1.2 (iii) are associative.

Exercise 1.4.

- (a) Prove that the natural numbers with addition and with multiplication form semigroups, while for the odd natural numbers, a semigroup arises only under multiplication.
- (b) Find other proper subsets of the natural numbers \mathbb{N} that form semigroups under addition or multiplication.

Exercise 1.5. Does the set \mathbb{N} of natural numbers under the operation of exponentiation,

$$n \circ m := n^m \quad (m, n \in \mathbb{N}),$$

form a semigroup?

Definition 1.6. A semigroup (H, \circ) is said to be *commutative*, or *abelian*, if for all elements $h_1, h_2 \in H$, we have

$$h_1 \circ h_2 = h_2 \circ h_1.$$

The term *abelian* is in honor of the Norwegian mathematician Niels Henrik Abel.

Example 1.7. The examples of semigroups (i) and (iii) above are both abelian. Example (ii) exhibits a semigroup that is in general nonabelian.

Exercise 1.8. Find two sets A_1 and A_2 such that $(\text{map}(A_1), \circ)$ is an abelian semigroup but $(\text{map}(A_2), \circ)$ is a nonabelian semigroup.

A modest generalization of the notion of semigroup leads to the concept of a monoid.

Definition 1.9. A *monoid* is a semigroup (H, \circ) that contains an *identity element* e with respect to the operation \circ , that is, an element such that

$$e \circ h = h = h \circ e$$

for every $h \in H$.

Lemma 1.10. The identity element e of a monoid (H, \circ) is uniquely determined.

Proof. Let e, e' be identity elements of the monoid (H, \circ) . By applying the identity element e , we obtain the equality

$$e \circ e' = e' = e' \circ e. \quad (3)$$

If we now bring the identity element e' into play, we obtain

$$e' \circ e = e = e \circ e'. \quad (4)$$

From equalities (3) and (4), we obtain at once the equality

$$e' = e' \circ e = e.$$

This proves the uniqueness of the identity element. \square

Remark 1.11. We can refine Definition 1.9 of a monoid by requiring only the existence of a *left identity element* e_ℓ (or *right identity element* e_r), which would satisfy the respective conditions

$$e_\ell \circ h = h \quad \text{and} \quad h \circ e_r = h$$

for all $h \in H$. However, it is easily shown that the left identity element is equal to the right identity element. We call such an element simply the identity element. With the preceding lemma, we can see that H has exactly one left identity element and one right identity element and that those two elements coincide.

Exercise 1.12. Let (H, \circ) be a semigroup and e_ℓ a left identity element and e_r a right identity element in H . Show that then $e_\ell = e_r$.

Example 1.13. The examples of semigroups from Example 1.2 are all examples of monoids:

- (i) The identity element of \mathbb{N} with respect to addition is 0; the identity element of \mathbb{N} with respect to multiplication is 1.
- (ii) The identity element of $(\text{map}(A), \circ)$ is the identity mapping $\text{id}_A : A \rightarrow A$, which maps every element $a \in A$ to itself.
- (iii) The identity element of \mathcal{R}_n with respect to \oplus is 0; the identity element of \mathcal{R}_n with respect to \odot is 1.

Exercise 1.14.

- (a) Show that the even natural numbers form a monoid under addition, but only a semigroup under multiplication.
- (b) Find other examples of semigroups that are not monoids.

2. Groups and Subgroups

We begin with the important definition of a group.

Definition 2.1. A monoid (G, \circ) with identity element e is called a *group* if for every $g \in G$, there exists an element $g' \in G$ such that

$$g' \circ g = e = g \circ g'.$$

Such an element g' is called an *inverse element* to g or simply an *inverse* of g .

Remark 2.2. In analogy to the uniqueness of the identity element of a monoid, one can show that the inverse g' of an element g of a group G is uniquely determined. We may therefore speak of *the* inverse g' of $g \in G$. The usual notation for the inverse g' of $g \in G$ is g^{-1} .

One can also refine Definition 2.1 of a group by requiring only the existence of a *left inverse* g'_ℓ (or a *right inverse* g'_r) for every $g \in G$, satisfying the respective conditions

$$g'_\ell \circ g = e \quad (\text{or } g \circ g'_r = e).$$

But as before, it can be shown that if there is a left inverse, then there is also a right inverse, and they are equal. Such an element is called simply an inverse element. We can then state that for every $g \in G$, there exists precisely one left inverse and one right inverse in G and that those two elements coincide.

Exercise 2.3.

- (a) Prove that the inverse g^{-1} of an element g of a group (G, \circ) is uniquely determined.
- (b) Let (G, \circ) be a group, $g \in G$, g'_ℓ a left inverse, and g'_r a right inverse of g . Show that $g'_\ell = g'_r$.

With the knowledge of the uniqueness of the identity element and inverses, we may state the definition of a group as follows.

Definition 2.4. A group (G, \circ) consists of a nonempty set G together with an associative operation \circ such that the following two properties are satisfied:

- (i) There exists a unique element $e \in G$ such that

$$e \circ g = g = g \circ e$$

for all $g \in G$. The element e is the *identity element* of G .

- (ii) For each $g \in G$, there exists a uniquely determined element $g^{-1} \in G$ such that

$$g^{-1} \circ g = e = g \circ g^{-1}.$$

The element g^{-1} is the *inverse element* to g .

Remark 2.5. For a group (G, \circ) with identity element e and $n \in \mathbb{N}$, we introduce the following useful exponential notation for the n -fold operation of an element $g \in G$ on itself:

$$g^n := \underbrace{g \circ \cdots \circ g}_{n \text{ times}} \text{ and } g^0 := e. \quad (5)$$

Exercise 2.6. Show that in the terminology of Remark 2.5, we have the following rules of calculation:

- (a) $(g^{-1})^{-1} = g$ for all $g \in G$.
- (b) $(g \circ h)^{-1} = h^{-1} \circ g^{-1}$ for all $g, h \in G$.
- (c) $g^n \circ g^m = g^{n+m}$ for all $g \in G$ and $n, m \in \mathbb{N}$.
- (d) $(g^n)^m = g^{n \cdot m}$ for all $g \in G$ and $n, m \in \mathbb{N}$.

Definition 2.7. A group (G, \circ) is called *commutative* or *abelian* if for all elements $g_1, g_2 \in G$, we have

$$g_1 \circ g_2 = g_2 \circ g_1.$$

Example 2.8. (i) $(G, \circ) = (\mathbb{N}, +)$ is not a group, since for no nonzero element $n \in \mathbb{N}$ does there exist a natural number n' that satisfies the equation $n' + n = 0 = n + n'$. That is, the nonzero natural numbers do not have (additive) inverses.

(ii) $(G, \circ) = (\mathcal{R}_n, \oplus)$ is a commutative group. If $a \in \mathcal{R}_n$, $a \neq 0$, then the inverse to a is given by the difference $n - a$, where we note that indeed, $n - a \in \mathcal{R}_n$.

The semigroup $(G, \circ) = (\mathcal{R}_n, \odot)$, on the other hand, is never a group, since the element 0 has no inverse. But even if we remove the zero element and consider the semigroup $(\mathcal{R}_n \setminus \{0\}, \odot)$, it is still not, in general, a group. If,

for example, we consider the case $n = 4$, then the element $2 \in \mathcal{R}_4$ has no inverse, for we have

$$2 \odot 0 = 0, \quad 2 \odot 1 = 2, \quad 2 \odot 2 = 0, \quad 2 \odot 3 = 2,$$

and so there is no $a \in \mathcal{R}_4$ such that $2 \odot a = 1$. If, however, we select a prime $p \in \mathbb{P}$, then it turns out that $(\mathcal{R}_p \setminus \{0\}, \odot)$ is a group.

(iii) Our next example of a group, the *dihedral group*, arises from geometry. Let $n \in \mathbb{N}$ be a nonzero natural number. For $n \geq 3$, let D_{2n} denote the set of all isometries of the Euclidean plane that map a regular n -gon to itself. The elements of D_{2n} are the rotations d_j through the angle $360^\circ \cdot j/n$ about the center M of the n -gon as well as the reflections s_j in the medians S_j when n is odd, and when n is even, the reflections s_j in the diagonals and the perpendicular bisectors S_j of the n -gon. In both the even and odd cases, we let the index j run from 0 to $n - 1$. Since the elements of D_{2n} are mappings, it makes sense to consider composition of mappings \circ as the operation. With this operation, D_{2n} becomes a monoid with identity element d_0 . Since each reflection $s_j \in D_{2n}$ can be written in the form $s_j = d_j \circ s_0$ with suitable numeration, we see that D_{2n} consists of the following $2n$ elements:

$$D_{2n} = \{d_0, d_1, \dots, d_{n-1}, d_0 \circ s_0, d_1 \circ s_0, \dots, d_{n-1} \circ s_0\}.$$

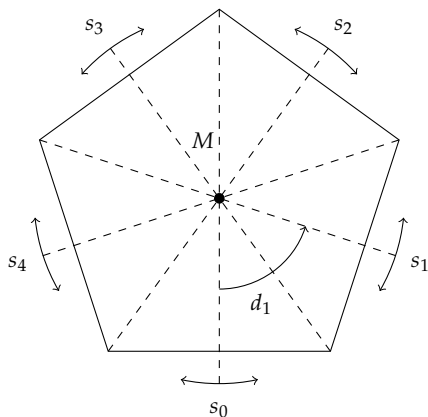


Fig. 1. Isometries of the regular pentagon.

Since every one of these elements obviously has an inverse (isometries of the plane are, after all, bijections), all the properties of a group are satisfied. We observe that the dihedral group (D_{2n}, \circ) for $n \geq 3$ is nonabelian, since, for example, $s_0 \circ d_1 = d_1^{-1} \circ s_0$.

For the cases $n = 1, 2$, we define the dihedral group analogously as follows:

$D_2 := \{d_0, s_0\}$ and $D_4 := \{d_0, d_1, s_0, d_1 \circ s_0\}$. We may interpret D_2 and D_4 as symmetry groups of the following 1-gon and 2-gon respectively:



Fig. 2. The 1-gon and 2-gon.

For $n = 1, 2$, the dihedral group (D_{2n}, \circ) is abelian.

(iv) As our last example, we consider a combinatorially based example of a group, the n th symmetric group:

$$S_n = \{\pi \mid \pi : \{1, \dots, n\} \longrightarrow \{1, \dots, n\} \text{ and } \pi \text{ is bijective}\}.$$

The elements of S_n can be written in the convenient form

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix},$$

where $\pi_j := \pi(j)$ for $1 \leq j \leq n$. For the associative operation on S_n , we again choose composition of mappings; that is, for $\pi, \sigma \in S_n$, we have

$$\pi \circ \sigma := \begin{pmatrix} 1 & 2 & \cdots & n \\ \tau_1 & \tau_2 & \cdots & \tau_n \end{pmatrix},$$

with $\tau_j := \pi(\sigma(j))$ for $1 \leq j \leq n$. The identity element is the identity permutation, given by the identity mapping on the set $\{1, \dots, n\}$. Furthermore, the existence of the inverse of a permutation is guaranteed by the fact that every bijective mapping $\pi : \{1, \dots, n\} \longrightarrow \{1, \dots, n\}$ has an inverse mapping π^{-1} . Under this operation, the set (S_n, \circ) forms a group, which for $n \geq 3$ is nonabelian.

Exercise 2.9. (Cayley tables). For a finite group, the result of the group operation on pairs of elements can be displayed in a *Cayley table*, named for the British mathematician Arthur Cayley, in which the elements of the group are listed in the first row and first column of a table, and the remaining fields are filled in with the result of the group operation. For example, the Cayley table for (\mathcal{R}_2, \oplus) is as follows:

\oplus	0	1
0	0	1
1	1	0

Fig. 3. Cayley table for the group (\mathcal{R}_2, \oplus)

Draw the Cayley tables for (\mathcal{R}_4, \oplus) , $(\mathcal{R}_5 \setminus \{0\}, \odot)$, (\mathcal{R}_6, \oplus) , (D_4, \circ) , and (D_6, \circ) , as well as for (S_2, \circ) and (S_3, \circ) . What similarities and differences do you notice?

Exercise 2.10.

- For the prime numbers $p = 3$ and $p = 5$, verify the assertion of Example 2.8 (ii) that $(\mathcal{R}_p \setminus \{0\}, \odot)$ is a group.
- Verify in detail the assertions of Example 2.8 (iii) regarding the dihedral group (D_{2n}, \circ) .
- Think about why the symmetric group (S_n, \circ) from Example 2.8 (iv) is nonabelian for all natural numbers $n \geq 3$.

Definition 2.11. Let (G, \circ) be a group. The cardinality of the set G underlying the group is called the *order* of G and is denoted by $|G|$. If the order of G is infinite, we write $|G| := \infty$.

Example 2.12. For the groups in Example 2.8 (ii) and (iii), we have

$$|\mathcal{R}_n| = n \quad \text{and} \quad |D_{2n}| = 2n.$$

Exercise 2.13. Show that for the symmetric group (S_n, \circ) , we have

$$|S_n| = n!.$$

Here $n!$ for $n \in \mathbb{N}$ is the factorial function, defined inductively as follows: $0! := 1$, $(n^*)! := n^* \cdot n!$.

Definition 2.14. A group (G, \circ) is said to be *cyclic* if there exists an element $g \in G$ such that

$$G = \{\dots, (g^{-1})^2, g^{-1}, g^0 = e, g^1 = g, g^2, \dots\}.$$

In such a case, we write $G = \langle g \rangle$ and say that g *generates* the group G .

Example 2.15. The group (\mathcal{R}_n, \oplus) is generated by the element 1, that is, we have $(\mathcal{R}_n, \oplus) = \langle 1 \rangle$, since every $a \in \mathcal{R}_n$ can be represented in the form

$$a = \underbrace{1 \oplus \dots \oplus 1}_{a \text{ times}}.$$

Remark 2.16. Let $G = \langle g \rangle$ be a cyclic group of order $n < \infty$. Then we have

$$G = \langle g \rangle = \{e, g, g^2, \dots, g^{n-1}\}.$$

This shows in particular that $g^n = e$, $g^{n+1} = g$, etc.

Definition 2.17. Let (G, \circ) be a group with identity element e and let g be an arbitrary element of G . The smallest nonzero natural number n such that $g^n = e$ is called the *order* of g and is denoted by $\text{ord}_G(g)$. If there is no such $n \in \mathbb{N}$, then the order of g is said to be infinite, that is, $\text{ord}_G(g) := \infty$.

If the group to which the order of g refers is clear from context, then we write simply $\text{ord}(g)$.

Example 2.18. We present here as an example the orders of the elements of the four-element group (\mathcal{R}_4, \oplus) :

$$\text{ord}(0) = 1, \quad \text{ord}(1) = 4, \quad \text{ord}(2) = 2, \quad \text{ord}(3) = 4.$$

Exercise 2.19. Determine the orders of all elements of the group S_3 .

Remark 2.20. Let $G = \langle g \rangle$ be a cyclic group of order $n < \infty$. Then $\text{ord}_G(g) = n$.

Definition 2.21. Let (G, \circ) be a group. A subset $U \subseteq G$ is called a *subgroup* of G if the restriction $\circ|_U$ of the operation \circ to U defines a group structure on U , that is, if $(U, \circ|_U)$ is itself a group. We express this relationship by writing $U \leq G$.

Example 2.22. Let m, n be natural numbers with $m \leq n$. Then the m th symmetric group S_m is a subgroup of the n th symmetric group S_n if we identify a permutation in S_m with the corresponding permutation of S_n that leaves $m + 1, \dots, n$ fixed. That is, $S_m \leq S_n$.

Exercise 2.23. Show that the rotations $\{d_0, \dots, d_{n-1}\}$ form a cyclic subgroup of the dihedral group D_{2n} .

Remark 2.24. Let (G, \circ) be a group, and U a subgroup of G . The identity element e of G is also the identity element of U . If h is an element of U , then its inverse in U is given by the inverse of h in G , that is, by h^{-1} , since

$$h \circ|_U h^{-1} = h \circ h^{-1} = e.$$

Lemma 2.25 (Subgroup criterion). Let (G, \circ) be a group and $U \subseteq G$ a nonempty subset. Then we have the equivalence

$$U \leq G \iff h_1 \circ h_2^{-1} \in U \quad \forall h_1, h_2 \in U.$$

Proof. (i) Suppose first that U is a subgroup of G . We must then show that for all $h_1, h_2 \in U$, we have the inclusion $h_1 \circ h_2^{-1} \in U$. But that is easy, since if $h_2 \in U$, then we also have $h_2^{-1} \in U$, and by applying the group operation to $h_1 \in U$, we at once obtain $h_1 \circ h_2^{-1} \in U$.

(ii) Now suppose that conversely, $h_1 \circ h_2^{-1} \in U$ for all $h_1, h_2 \in U$. Since U is nonempty, there is at least one element $h \in U$, for which we then have

$e = h \circ h^{-1} \in U$. That is, U contains the identity element. If h' is an arbitrary element of U , we see that

$$h'^{-1} = e \circ h'^{-1} \in U.$$

That is, $h' \in U$ implies that $h'^{-1} \in U$. Finally, let h_1 and h_2 be arbitrary elements of U . We must convince ourselves that the element $h_1 \circ h_2$ is also in U . We recall that $h_2 \in U$ implies $h_2^{-1} \in U$. Using rule (a) from Exercise 2.6, we obtain

$$h_1 \circ h_2 = h_1 \circ (h_2^{-1})^{-1} \in U.$$

We conclude that \circ is an associative operation defined on U and that (U, \circ) satisfies all the group axioms. This completes the proof that U is a subgroup of G . \square

Exercise 2.26. Find all subgroups of the group S_3 . Which of these are cyclic groups?

3. Group Homomorphisms

In this section, we are going to compare groups using mappings that respect the group operation. The first thing, then, is to explain what is meant by *preserving* the group operation, or group structure.

Definition 3.1. Let (G, \circ_G) and (H, \circ_H) be groups. A mapping

$$f : (G, \circ_G) \longrightarrow (H, \circ_H)$$

is called a *group homomorphism* if for all $g_1, g_2 \in G$, we have the equality

$$f(g_1 \circ_G g_2) = f(g_1) \circ_H f(g_2).$$

The significance of a homomorphism is, then, that the image under f of the composition of two elements g_1 and g_2 in G is equal to the composition of the images under f of g_1 and g_2 in H . We sometimes say that the mapping f *preserves the group structure*.

A bijective (that is, both injective and surjective) group homomorphism is called a *group isomorphism*. If $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ is a group isomorphism, we say that the groups G and H are *isomorphic*, and we write $G \cong H$.

Example 3.2. Consider the dihedral group $G = D_6$ and the symmetric group $H = S_3$. The dihedral group D_6 consists of all symmetries of an equilateral triangle \triangle . Let us denote the vertices of \triangle in counterclockwise order by the natural numbers 1, 2, 3. If we choose one of the symmetry mappings $g \in D_6$ and allow it to act on \triangle , the result is a permutation π of the set $\{1, 2, 3\}$. The assignment $g \mapsto \pi$ thereby induces a mapping

$$f : D_6 \longrightarrow S_3.$$

If we consider all possible compositions of symmetries and their images under f and compare them with the corresponding compositions of permutations, we see that f is a group homomorphism.

Exercise 3.3. Is this mapping also a group isomorphism?

Definition 3.4. Let (G, \circ_G) be a group with identity element e_G , and let (H, \circ_H) be a group with identity element e_H . Furthermore, let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a group homomorphism. Then

$$\ker(f) := \{g \in G \mid f(g) = e_H\}$$

is called the *kernel* of f , and

$$\text{im}(f) := \{h \in H \mid \exists g \in G : h = f(g)\}$$

is called the *image* of f .

Exercise 3.5. Let D_{2n} be the dihedral group from Example 2.8(iii). In that example, we noted that every element can be expressed uniquely in the form $d_j \circ s_0^k$ with $j \in \{0, \dots, n-1\}$ and $k \in \{0, 1\}$. Show that the mapping $\text{sgn} : (D_{2n}, \circ) \longrightarrow (\mathcal{R}_2, \oplus)$, given by the assignment $d_j \circ s_0^k \mapsto k$, is a group homomorphism, and determine the kernel and image of sgn .

Lemma 3.6. Let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a group homomorphism. Then we have the following:

- (i) f is injective if and only if $\ker(f) = \{e_G\}$.
- (ii) f is surjective if and only if $\text{im}(f) = H$.

Proof. (i) By definition, the mapping f is injective if and only if for all $g_1, g_2 \in G$,

$$f(g_1) = f(g_2) \tag{6}$$

implies that g_1 and g_2 are equal. We therefore take equality (6) and transform it by means of the group homomorphism property of f into the equivalent form

$$f(g_1) \circ_H (f(g_2))^{-1} = e_H \iff f(g_1) \circ_H f(g_2^{-1}) = e_H.$$

Applying again the group homomorphism property of f yields $f(g_1 \circ_G g_2^{-1}) = e_H$, that is, $g_1 \circ_G g_2^{-1} \in \ker(f)$. Finally, the equivalence

$$g_1 \circ_G g_2^{-1} = e_G \iff g_1 = g_2$$

shows that we have $\ker(f) = \{e_G\}$ if and only if $g_1 = g_2$, that is, if and only if f is injective.

(ii) The proof of this assertion is obvious, since the surjectivity of f means precisely that every element of H is the image of some element of G under the mapping f . \square

Exercise 3.7. Let $f : (G, \circ) \longrightarrow (H, \circ)$ be a group homomorphism and assume that $|G| < \infty$. Prove the equivalence

$$\ker(f) = \{e_G\} \iff f \text{ is a group isomorphism.}$$

Exercise 3.8. Let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a group homomorphism. Show that for every element $g \in G$, we have $\text{ord}_G(g) \geq \text{ord}_H(f(g))$.

Exercise 3.9. Does there exist a group isomorphism between D_{24} and S_4 ?

Lemma 3.10. Let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a group homomorphism. Then $\ker(f)$ is a subgroup of G , and $\text{im}(f)$ is a subgroup of H .

Proof. We begin with the proof that $\ker(f)$ is a subgroup of G . We first observe that because we have $f(e_G) = e_H$, that is, $e_G \in \ker(f)$, the kernel of f is nonempty. We now apply the subgroup criterion (Lemma 2.25). To this end, we choose $g_1, g_2 \in \ker(f)$ and must show that $g_1 \circ_G g_2^{-1} \in \ker(f)$. But this follows easily from the homomorphism property of f :

$$f(g_1 \circ_G g_2^{-1}) = f(g_1) \circ_H f(g_2^{-1}) = e_H \circ_H (f(g_2))^{-1} = e_H \circ_H e_H^{-1} = e_H.$$

To prove the subgroup property of $\text{im}(f)$, we proceed analogously. Again, since $e_H = f(e_G)$, that is, $e_H \in \text{im}(f)$, the image of f is nonempty. We again make use of the subgroup criterion and must establish for $h_1, h_2 \in \text{im}(f)$ the relationship $h_1 \circ_H h_2^{-1} \in \text{im}(f)$. Since $h_1, h_2 \in \text{im}(f)$, there exist $g_1, g_2 \in G$ such that $h_1 = f(g_1)$ and $h_2 = f(g_2)$. Again using the homomorphism property of f yields

$$h_1 \circ_H h_2^{-1} = f(g_1) \circ_H (f(g_2))^{-1} = f(g_1) \circ_H f(g_2^{-1}) = f(g_1 \circ_G g_2^{-1});$$

that is, the element $h_1 \circ_H h_2^{-1}$ is the image of the element $g_1 \circ_G g_2^{-1}$. This completes the proof of the lemma. \square

Exercise 3.11.

- (a) Find all group homomorphisms $f : (\mathcal{R}_4, \oplus) \longrightarrow (\mathcal{R}_4, \oplus)$.
- (b) Let p be a prime number and $n \in \mathbb{N}$ a natural number that is not divisible by p . Find all group homomorphisms $g : (\mathcal{R}_p, \oplus) \longrightarrow (\mathcal{R}_n, \oplus)$. Determine the image and kernel of each homomorphism.

4. Cosets and Normal Subgroups

Before we introduce the notion of a coset (with respect to a subgroup), we recall the definition of an equivalence relation.

Definition 4.1. Let M be a set. A (binary) relation \sim on M is called an *equivalence relation* if the following three conditions are satisfied:

- (i) The relation \sim is *reflexive*, that is, for all $m \in M$, we have $m \sim m$.
- (ii) The relation \sim is *symmetric*, that is, for all $m_1, m_2 \in M$ such that $m_1 \sim m_2$, we have also $m_2 \sim m_1$.
- (iii) The relation \sim is *transitive*, that is, for all $m_1, m_2, m_3 \in M$ such that $m_1 \sim m_2$ and $m_2 \sim m_3$, we have also $m_1 \sim m_3$.

Example 4.2. The equality “=” of elements of a set defines an equivalence relation.

Exercise 4.3.

- (a) Verify the assertion of Example 4.2.
- (b) Is the order relation \leq on \mathbb{N} an equivalence relation?
- (c) Consider a relation \sim on the set of natural numbers \mathbb{N} whereby $m \sim n$ if m is a power of n or n is a power of m . Determine whether \sim is an equivalence relation.

Remark 4.4. Let M be a set equipped with an equivalence relation \sim . For each $m \in M$, we can construct the set

$$M_m := \{m' \in M \mid m' \sim m\}.$$

The set M_m is called the *equivalence class* of m .

Lemma 4.5. Let M be a set equipped with an equivalence relation \sim . Then we have the following:

- (i) Two equivalence classes in M are either disjoint or identical.
- (ii) The set M is the disjoint union of its equivalence classes. We indicate this by writing

$$M = \dot{\bigcup}_{m \in I} M_m,$$

where $I \subseteq M$ is a subset containing exactly one representative from each equivalence class.

Proof. (i) Let $m_1, m_2 \in M$ be such that $M_{m_1} \cap M_{m_2} \neq \emptyset$, where \emptyset is the standard notation for the empty set. We must show that $M_{m_1} = M_{m_2}$. Since $M_{m_1} \cap M_{m_2} \neq \emptyset$, there exists $m \in M_{m_1} \cap M_{m_2}$; that is, we have $m \sim m_1$ and $m \sim m_2$, and therefore, by the symmetry and transitivity of the equivalence relation \sim , we have $m_1 \sim m_2$, whence we have $m_1 \in M_{m_2}$. It follows by another application of transitivity that we likewise have $m' \in M_{m_2}$

for all $m' \in M_{m_1}$. We see, then, that $M_{m_1} \subseteq M_{m_2}$. Interchanging the roles of the equivalence classes M_{m_1} and M_{m_2} , we obtain the reverse inclusion $M_{m_2} \subseteq M_{m_1}$, from which follows the equality $M_{m_1} = M_{m_2}$.

(ii) To prove the second part of the assertion, we begin with the case that M is a finite set. In this case, we can proceed constructively. If M is empty, then there is nothing to prove. Otherwise, there exists $m_1 \in M$ with its equivalence class M_{m_1} . The set-theoretic difference $M \setminus M_{m_1}$ is now either empty, that is, $M = M_{m_1}$, or there exists $m_2 \in M \setminus M_{m_1}$ with equivalence class M_{m_2} . We have now the two alternatives

$$M = M_{m_1} \dot{\cup} M_{m_2} \quad \text{and} \quad \exists m_3 \in M \setminus (M_{m_1} \dot{\cup} M_{m_2}).$$

Since the set M is finite, this process must end after finitely many steps, say k steps, and we obtain M as the disjoint union

$$M = \bigcup_{j=1}^k M_{m_j}.$$

Now that we have illustrated the proof in the case of finite sets, let us turn our attention to the general situation. Since the equivalence class M_m associated with $m \in M$ contains the element m , it is clear that M is the union of all its equivalence classes. That is,

$$M = \bigcup_{m \in M} M_m.$$

This union, however, is not in general disjoint. By selecting a unique representative of each equivalence class, we obtain a subset $I \subseteq M$ such that for each $m \in I$, the associated equivalence class M_m in the above union appears exactly once. The subset I is called a complete set of equivalence class representatives. We thereby obtain the representation of M as the disjoint union

$$M = \dot{\bigcup}_{m \in I} M_m,$$

as asserted. □

Exercise 4.6. Describe the equivalence classes of the equality relation “=” from Example 4.2. Come up with other equivalence relations and determine the associated equivalence classes.

We now introduce a particular equivalence relation on a group induced by a subgroup.

Remark 4.7. Let (G, \circ) be a group, and $U \leq G$ a subgroup. We define on G the relation

$$g_1 \sim g_2 \iff g_1^{-1} \circ g_2 \in U \quad (g_1, g_2 \in G).$$

We assert that this defines an equivalence relation on G . The reflexivity $g \sim g$ is immediate from the fact that $g^{-1} \circ g = e \in U$. If $g_1 \sim g_2$, whence $g_1^{-1} \circ g_2 \in U$, it follows by taking inverses that

$$U \ni (g_1^{-1} \circ g_2)^{-1} = g_2^{-1} \circ g_1.$$

That is, $g_2 \sim g_1$, which proves symmetry. Finally, if we have $g_1 \sim g_2$ and $g_2 \sim g_3$, whence $g_1^{-1} \circ g_2 \in U$ and $g_2^{-1} \circ g_3 \in U$, it follows by composition that

$$U \ni (g_1^{-1} \circ g_2) \circ (g_2^{-1} \circ g_3) = g_1^{-1} \circ g_3,$$

that is, $g_1 \sim g_3$, which establishes transitivity.

Definition 4.8. Let (G, \circ) be a group, $U \leq G$ a subgroup, and \sim the equivalence relation from Remark 4.7. We call the equivalence class of $g \in G$, that is, the set of group elements

$$\{g' \in G \mid g' \sim g\},$$

the *left coset* of g with respect to the subgroup U . From the equivalence

$$g' \sim g \iff g^{-1} \circ g' \in U \iff \exists h \in U : g' = g \circ h,$$

we obtain

$$\{g' \in G \mid g' \sim g\} = \{g \circ h \mid h \in U\}.$$

We may therefore denote the left coset of g with respect to U simply by $g \circ U$.

Remark 4.9. Let (G, \circ) be a group, $U \leq G$ a subgroup, and \sim the equivalence relation from Remark 4.7. Then using Lemma 4.5, we obtain a decomposition of G into disjoint left cosets; that is,

$$G = \dot{\bigcup}_{g \in I} g \circ U,$$

where $I \subseteq G$ is a complete set of representatives of all left cosets with respect to U .

Definition 4.10. Let (G, \circ) be a group and $U \leq G$ a subgroup. We denote by G/U the set of all left cosets of elements of G with respect to U , that is,

$$G/U = \{g \circ U \mid g \in I\},$$

where $I \subseteq G$ is a complete set of representatives of all left cosets with respect to U .

Exercise 4.11. Let m, n be natural numbers with $1 \leq m \leq n$. Find a complete set of representatives of the set of left cosets S_n/S_m .

Exercise 4.12. From among the subgroups of S_3 determined in Exercise 2.26, choose a subgroup of order two and determine all left cosets of S_3 with respect to this subgroup.

Lemma 4.13. Let (G, \circ) be a group, and $U \leq G$ a subgroup. All left cosets of G with respect to U have the same order as the subgroup U .

Proof. Let $g \circ U$ be the left coset of g with respect to U , and consider the mapping

$$\varphi : g \circ U \longrightarrow U,$$

given by $g \circ h \mapsto h$ ($h \in U$). The assignment $h \mapsto g \circ h$ clearly induces the inverse mapping to φ , namely φ^{-1} . We see, then, that φ is bijective, from which it follows that $g \circ U$ and U have the same order. That is, we have the equality

$$|g \circ U| = |U|,$$

as asserted. □

Theorem 4.14 (Lagrange's theorem). Let (G, \circ) be a finite group (that is, $|G| < \infty$), and let $U \leq G$ be a subgroup. Then the order of U divides the order of G , that is, $|U| \mid |G|$.

Proof. Since the group G is finite, it can be decomposed into finitely many left cosets with respect to U . That is, we have a disjoint decomposition of the form

$$G = (g_1 \circ U) \dot{\cup} \cdots \dot{\cup} (g_k \circ U).$$

Since the left cosets $g_j \circ U$ ($j = 1, \dots, k$) are mutually disjoint and each of their orders is equal to $|U|$ by Lemma 4.13, we obtain

$$|G| = \sum_{j=1}^k |g_j \circ U| = k \cdot |U|.$$

This completes the proof of the theorem. □

Exercise 4.15.

- (a) Derive from Lagrange's theorem the fact that in a finite group, the order of each element is a divisor of the order of the group.
- (b) Conclude from part (a) that a group whose order is a prime number must be cyclic.
- (c) Determine all possible groups of orders 4 and 6 up to isomorphism.

Definition 4.16. Let (G, \circ) be a group, and $U \leq G$ a subgroup. The order of G/U is called the *index of U in G* and is denoted by $[G : U]$.

Remark 4.17. If (G, \circ) is a finite group and $U \leq G$ a subgroup, it follows from the proof of Lagrange's theorem that the order of G is equal to the product of the order of U and the index of U in G . That is, we have

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

In analogy to the left cosets, we can, of course, construct the set of right cosets.

Remark 4.18. Let (G, \circ) be a group, and $U \leq G$ a subgroup. We define on G the additional relation

$$g_1 \sim_r g_2 \iff g_1 \circ g_2^{-1} \in U \quad (g_1, g_2 \in G).$$

We leave it as an exercise to the reader to show that this defines an equivalence relation on G . The equivalence class of $g \in G$ is called the *right coset of g with respect to U* . This leads to the following:

$$\{g' \in G \mid g' \sim_r g\} = \{h \circ g \mid h \in U\} =: U \circ g.$$

We have thus obtained a decomposition of G into disjoint right cosets; that is,

$$G = \bigcup_{g \in I_r} U \circ g,$$

where $I_r \subseteq G$ is a complete system of right coset representatives with respect to U .

We denote the set of right cosets with respect to U by $U \backslash G$. Just as in the case of left cosets, all the right cosets of G with respect to U have the same order as the subgroup U .

Finally, it is easy to verify that by associating the left coset $g \circ U$ with the right coset $U \circ g^{-1}$, we induce a bijection between the sets G/U and $U \backslash G$. That is, we have

$$|G/U| = [G : U] = |U \backslash G|.$$

If the group G is abelian, then the left and right cosets coincide.

Exercise 4.19. Solve Exercises 4.11 and 4.12 for right cosets.

Definition 4.20. Let (G, \circ) be a group. A subgroup N of G is said to be a *normal subgroup* if all left and right cosets with respect to N coincide, that is, if for all $g \in G$, we have $g \circ N = N \circ g$.

Since left and right cosets with respect to a normal subgroup N coincide, we speak in this case simply of *cosets*. If $N \leq G$ is a normal subgroup, then we indicate this fact by writing $N \trianglelefteq G$.

Exercise 4.21. Is the subgroup chosen in Exercise 4.12 normal?

Remark 4.22. The following is equivalent to the definition above: A subgroup N of G is normal if and only if for every $g \in G$, we have

$$g \circ N \circ g^{-1} = N,$$

where

$$g \circ N \circ g^{-1} = \{g' \in G \mid g' = g \circ h \circ g^{-1} \text{ with } h \in N\}.$$

We have yet another equivalent description of a normal subgroup: a subgroup N of G is normal if and only if for all $g \in G$ and $h \in N$, we have $g \circ h \circ g^{-1} \in N$. We can see that this definition is equivalent to the previous one: We note first that we clearly have $g \circ N \circ g^{-1} \subseteq N$ for all $g \in G$. To prove the reverse inclusion, we observe that from $g \circ h \circ g^{-1} \in N$ for all $g \in G, h \in N$, we have in particular that $g^{-1} \circ h \circ g \in N$ for all $g \in G, h \in N$. From this we conclude that $g^{-1} \circ N \circ g \subseteq N$ for all $g \in G$. By operating on this relation on the left by g and on the right by g^{-1} , we obtain

$$N = g \circ (g^{-1} \circ N \circ g) \circ g^{-1} \subseteq g \circ N \circ g^{-1},$$

which is precisely the desired reverse inclusion. Therefore, we have indeed the equality $g \circ N \circ g^{-1} = N$ for all $g \in G$.

Example 4.23. We now consider the example of a normal subgroup of the symmetric group S_3 . The reader will recall that S_3 is given by the six permutations

$$S_3 = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\},$$

where

$$\begin{aligned} \pi_1 &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, & \pi_2 &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, & \pi_3 &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \\ \pi_4 &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, & \pi_5 &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, & \pi_6 &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}. \end{aligned}$$

The three permutations π_1, π_2, π_3 form a cyclic subgroup of order 3, denoted by $A_3 = \langle \pi_2 \rangle$ and called the *alternating group of degree 3*. We shall now prove that A_3 is a normal subgroup of S_3 . For $j = 1, 2, 3$, we have the obvious equality

$$\pi_j \circ A_3 = A_3 = A_3 \circ \pi_j.$$

An explicit calculation with the element π_4 shows that

$$\begin{aligned} \pi_4 \circ A_3 &= \{\pi_4 \circ \pi_1, \pi_4 \circ \pi_2, \pi_4 \circ \pi_3\} = \{\pi_4, \pi_5, \pi_6\}, \\ A_3 \circ \pi_4 &= \{\pi_1 \circ \pi_4, \pi_2 \circ \pi_4, \pi_3 \circ \pi_4\} = \{\pi_4, \pi_6, \pi_5\}, \end{aligned}$$

which establishes the equality $\pi_4 \circ A_3 = A_3 \circ \pi_4$. One can perform a similar calculation for $j = 5, 6$:

$$\pi_j \circ A_3 = A_3 \circ \pi_j,$$

which proves the normality of A_3 . Our calculations have shown furthermore that the set of (left) cosets with respect to A_3 is given by

$$S_3/A_3 = \{A_3, \pi_4 \circ A_3\}.$$

In particular, we see that

$$[S_3 : A_3] = |S_3|/|A_3| = \frac{6}{3} = 2.$$

Exercise 4.24. Let G be a group and $H \leq G$ a subgroup of index 2.

(a) Show that H is a normal subgroup of G .

(b) Give a surjective group homomorphism from G to the group (\mathcal{R}_2, \oplus) .

Lemma 4.25. Let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a group homomorphism. Then the kernel $\ker(f)$ of f is a normal subgroup of G .

Proof. For simplicity of notation, we shall write simply \circ in place of both \circ_G and \circ_H .

By Lemma 3.10, $\ker(f)$ is a subgroup of G . It remains to prove the normality property for $\ker(f)$, namely that

$$g \circ h \circ g^{-1} \in \ker(f)$$

for all $g \in G$ and $h \in \ker(f)$. So let $g \in G$ and $h \in \ker(f)$ be arbitrary elements. We observe that $f(h) = e_H$. Using the homomorphism property of f , we obtain

$$\begin{aligned} f(g \circ h \circ g^{-1}) &= f(g) \circ f(h) \circ f(g^{-1}) = f(g) \circ e_H \circ (f(g))^{-1} \\ &= f(g) \circ f(g)^{-1} = e_H. \end{aligned}$$

We have therefore $g \circ h \circ g^{-1} \in \ker(f)$, and the lemma is proved. \square

Exercise 4.26. Let $f : (S_3, \circ) \longrightarrow (\mathcal{R}_3, \oplus)$ be a group homomorphism. Show that we must have $f(\pi) = 0$ for all $\pi \in S_3$.

5. Quotient Groups and the Homomorphism Theorem

We shall now show how we can provide, in a natural way, the set G/N of (left) cosets of a group G with respect to a normal subgroup N with a group structure. As a rule, the structure of the group G/N will be in some respect

simpler than the structure of the group G . Studying the group G/N provides information about the structure of the group G .

Definition 5.1. Let (G, \circ) be a group, and $N \trianglelefteq G$ a normal subgroup. We define an operation \bullet on the set of (left) cosets with respect to N as follows:

$$(g_1 \circ N) \bullet (g_2 \circ N) := (g_1 \circ g_2) \circ N \quad (g_1, g_2 \in G). \quad (7)$$

This definition appears to depend on the choice of representatives g_1 and g_2 for the cosets $g_1 \circ N$ and $g_2 \circ N$. We shall show, however, in the following lemma that the operation \bullet is in fact independent of the choice of representatives.

Lemma 5.2. Let (G, \circ) be a group, and $N \trianglelefteq G$ a normal subgroup. Then the operation \bullet defined on G/N in Definition 5.1 is well defined.

Proof. Let g_1, g'_1 and g_2, g'_2 be representatives of the respective cosets $g_1 \circ N$ and $g_2 \circ N$. To prove that the operation (7) is independent of the choice of representatives, we must prove the equality

$$(g_1 \circ g_2) \circ N = (g'_1 \circ g'_2) \circ N.$$

Since $g'_1 \in g_1 \circ N$, there exists $h_1 \in N$ such that $g'_1 = g_1 \circ h_1$; analogously, we obtain $g'_2 = g_2 \circ h_2$ for some $h_2 \in N$. We now calculate, taking into account the associativity of \circ ,

$$\begin{aligned} (g'_1 \circ g'_2) \circ N &= ((g_1 \circ h_1) \circ (g_2 \circ h_2)) \circ N = (g_1 \circ h_1 \circ g_2) \circ (h_2 \circ N) \\ &= (g_1 \circ (h_1 \circ g_2)) \circ N, \end{aligned}$$

where in the last step, we used the equality $h_2 \circ N = N$, which holds because we have $h_2 \in N$. Since N is normal in G , there exists $h'_1 \in N$ such that $h_1 \circ g_2 = g_2 \circ h'_1$. Substituting this in the previous equation yields, as asserted,

$$(g'_1 \circ g'_2) \circ N = (g_1 \circ (g_2 \circ h'_1)) \circ N = (g_1 \circ g_2) \circ N;$$

here we have again used the associativity of \circ and the equality $h'_1 \circ N = N$. This completes the proof of the lemma. \square

With the help of Lemma 5.2, we now have a well-defined operation, namely \bullet , on the set G/N . The following proposition asserts that $(G/N, \bullet)$ is in fact a group.

Proposition 5.3. Let (G, \circ) be a group, and $N \trianglelefteq G$ a normal subgroup. The set G/N of (left) cosets of G with respect to N together with the operation \bullet forms a group.

Proof. We begin by establishing that the set G/N is nonempty, which can be seen from the fact that it contains the coset $e_G \circ N = N$, that is, the element N . The associativity of the operation \bullet follows at once from that of the operation \circ on the group G . Namely, using the definition of \bullet and Lemma 5.2, we obtain

$$\begin{aligned} & ((g_1 \circ N) \bullet (g_2 \circ N)) \bullet (g_3 \circ N) \\ &= ((g_1 \circ g_2) \circ N) \bullet (g_3 \circ N) = ((g_1 \circ g_2) \circ g_3) \circ N \\ &= (g_1 \circ (g_2 \circ g_3)) \circ N = (g_1 \circ N) \bullet ((g_2 \circ g_3) \circ N) \\ &= (g_1 \circ N) \bullet ((g_2 \circ N) \bullet (g_3 \circ N)) \end{aligned}$$

for all $g_1, g_2, g_3 \in G$. The identity element of G/N is given by N . Indeed, for every coset $g \circ N \in G/N$, we have

$$\begin{aligned} N \bullet (g \circ N) &= (e_G \circ N) \bullet (g \circ N) = (e_G \circ g) \circ N = g \circ N, \\ (g \circ N) \bullet N &= (g \circ N) \bullet (e_G \circ N) = (g \circ e_G) \circ N = g \circ N. \end{aligned}$$

Finally, the inverse element to $g \circ N$ is given by the coset $g^{-1} \circ N$, for we have

$$\begin{aligned} (g^{-1} \circ N) \bullet (g \circ N) &= (g^{-1} \circ g) \circ N = e_G \circ N = N, \\ (g \circ N) \bullet (g^{-1} \circ N) &= (g \circ g^{-1}) \circ N = e_G \circ N = N. \end{aligned}$$

Thus $(G/N, \bullet)$ satisfies all the properties of a group, and the lemma is proved. \square

Definition 5.4. Let (G, \circ) be a group, and $N \trianglelefteq G$ a normal subgroup. The group $(G/N, \bullet)$ is called the *quotient group* of G by the normal subgroup N .

Example 5.5. (i) In an abelian group G , every subgroup H is normal. Therefore, we can form the quotient group $(G/H, \bullet)$ for every subgroup H of G . Each such quotient group is abelian.

(ii) In Example 4.23, we proved that the alternating group A_3 is a normal subgroup of the symmetric group S_3 . We may therefore form the quotient group S_3/A_3 , which (in the notation of Example 4.23) consists of the two elements $e := A_3$ and $g := \pi_4 \circ A_3$. The element e is the identity element in S_3/A_3 , and the element g satisfies the relation $g \bullet g = e$. We may therefore identify the quotient group S_3/A_3 with the familiar group (\mathcal{R}_2, \oplus) , which consists of the elements 0 and 1, by mapping the element e to 0 and the element g to 1. It is easy to see that this identification is a bijective group homomorphism from S_3/A_3 to \mathcal{R}_2 . We have therefore the group isomorphism

$$(S_3/A_3, \bullet) \cong (\mathcal{R}_2, \oplus).$$

Remark 5.6. Let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a group homomorphism. Lemma 4.25 asserts that $\ker(f)$ is a normal subgroup of G . We may therefore form the quotient group $(G/\ker(f), \bullet)$. We now define the mapping

$$\pi : (G, \circ_G) \longrightarrow (G/\ker(f), \bullet)$$

via $g \mapsto g \circ_G \ker(f)$. The definition of the operation \bullet now shows that

$$\begin{aligned} \pi(g_1 \circ_G g_2) &= (g_1 \circ_G g_2) \circ_G \ker(f) = (g_1 \circ_G \ker(f)) \bullet (g_2 \circ_G \ker(f)) \\ &= \pi(g_1) \bullet \pi(g_2); \end{aligned}$$

that is, the mapping π is a group homomorphism, and it is surjective. The homomorphism π is called the *canonical group homomorphism*.

Theorem 5.7 (Homomorphism theorem for groups). *Let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a group homomorphism. Then f induces a uniquely determined injective group homomorphism*

$$\bar{f} : (G/\ker(f), \bullet) \longrightarrow (H, \circ_H)$$

such that $\bar{f}(g \circ_G \ker(f)) = f(g)$ for all $g \in G$. The statement of the theorem can be illustrated by saying that the diagram

$$\begin{array}{ccc} (G, \circ_G) & & \\ \pi \downarrow & \searrow f & \\ (G/\ker(f), \bullet) & \xrightarrow{\exists! \bar{f}} & (H, \circ_H) \end{array}$$

is commutative, that is, that we obtain the same result by applying the mapping f directly or by first applying π and then the mapping \bar{f} .

Proof. To simplify notation, we define $N := \ker(f)$, and furthermore, we shall write simply \circ in place of \circ_G and \circ_H . By Lemma 4.25, N is a normal subgroup of G . We thereby obtain the quotient group $(G/N, \bullet)$. We now define a mapping \bar{f} from $(G/N, \bullet)$ to (H, \circ_H) as follows:

$$\bar{f}(g \circ N) := f(g) \quad (g \in G).$$

Since we defined \bar{f} in terms of a particular representative g of the coset $g \circ N$, we must show that \bar{f} is well defined. To this end, let $g' \in G$ be an arbitrary representative of the coset $g \circ N$; that is, there exists $h \in N$ such that $g' = g \circ h$. We then obtain

$$f(g') = f(g \circ h) = f(g) \circ f(h) = f(g) \circ e_H = f(g),$$

which shows that the definition of \bar{f} is independent of the choice of representative of $g \circ N$.

In a further step, we show that \bar{f} is a group homomorphism. Choose two arbitrary cosets $g_1 \circ N$ and $g_2 \circ N$ in G/N , and using the definition of \bar{f} and the homomorphism f , calculate

$$\begin{aligned}\bar{f}((g_1 \circ N) \bullet (g_2 \circ N)) &= \bar{f}((g_1 \circ g_2) \circ N) = f(g_1 \circ g_2) = f(g_1) \circ f(g_2) \\ &= \bar{f}(g_1 \circ N) \circ \bar{f}(g_2 \circ N).\end{aligned}$$

This shows that \bar{f} is in fact a homomorphism.

In a third step, we show the injectivity of \bar{f} . Let $g_1 \circ N, g_2 \circ N \in G/N$ be such that $\bar{f}(g_1 \circ N) = \bar{f}(g_2 \circ N)$. We have to show that $g_1 \circ N = g_2 \circ N$. By definition, this proposed equality is equivalent to the equality $f(g_1) = f(g_2)$. If we apply $f(g_1)^{-1}$ to both sides of this equality from the left, we obtain

$$e_H = f(g_1)^{-1} \circ f(g_1) = f(g_1)^{-1} \circ f(g_2) = f(g_1^{-1} \circ g_2);$$

that is, we have $g_1^{-1} \circ g_2 \in \ker(f) = N$. This yields at once that g_2 is an element of the coset $g_1 \circ N$, that is, $g_2 \sim g_1$. We have, therefore, the equality

$$g_1 \circ N = g_2 \circ N,$$

as asserted. Putting all of this together, we have shown that

$$\bar{f} : (G/\ker(f), \bullet) \longrightarrow (H, \circ_H)$$

is a well-defined injective group homomorphism. It remains to prove the uniqueness of \bar{f} such that $\bar{f}(g \circ \ker(f)) = f(g)$ ($g \in G$). Let

$$\tilde{f} : (G/\ker(f), \bullet) \longrightarrow (H, \circ_H)$$

be another injective group homomorphism such that $\tilde{f}(g \circ \ker(f)) = f(g)$ ($g \in G$). Then we have

$$\tilde{f}(g \circ \ker(f)) = f(g) = \bar{f}(g \circ \ker(f)) \quad (g \in G),$$

which means precisely that the action of \tilde{f} is identical to the action of \bar{f} on $(G/\ker(f), \bullet)$. That is, we have $\tilde{f} = \bar{f}$, which proves the uniqueness of \bar{f} . This completes the proof of the homomorphism theorem for groups. \square

Corollary 5.8. *Let $f : (G, \circ_G) \longrightarrow (H, \circ_H)$ be a surjective group homomorphism. Then f induces a uniquely determined group isomorphism*

$$\bar{f} : (G/\ker(f), \bullet) \cong (H, \circ_H)$$

such that $\bar{f}(g \circ_G \ker(f)) = f(g)$ for all $g \in G$. \square

Example 5.9. We consider the symmetric group S_n and recall from linear algebra that every permutation π can be written as a composition of transpositions (i.e., permutations that interchange two elements and leave the others fixed) and that while such a representation is not unique, the number of transpositions that occur in the representation of a given permutation is always even or always odd, and depending on which it is, we speak of a permutation as itself being either even or odd. We may therefore define the mapping

$$f : (S_n, \circ) \longrightarrow (\mathcal{R}_2, \oplus)$$

by sending π to 0 if the permutation is even, and to 1 if it is odd. It is easily verified that f is a surjective group homomorphism. The kernel $\ker(f)$ of f consists of the even permutations, that is, those that can be represented by an even number of transpositions. We call this subgroup the alternating group of degree n and denote it by A_n . By Corollary 5.8, we obtain the group isomorphism

$$(S_n / A_n, \bullet) \cong (\mathcal{R}_2, \oplus).$$

Exercise 5.10. Generalize the above discussion to the case of Exercise 4.24. That is, construct a group isomorphism

$$(G/H, \bullet) \cong (\mathcal{R}_2, \oplus)$$

for a subgroup $H \leq G$ of index 2.

From the homomorphism theorem for groups, one can deduce a number of additional isomorphisms between groups. Here is a typical example.

Exercise 5.11. Let G be a group, and $H, K \trianglelefteq G$ normal subgroups in G such that $K \subseteq H$. Show that K is normal in H , and we have the isomorphism

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

6. Construction of Groups from Regular Semigroups

In Remark 1.26 of Chapter I, we noted the bothersome fact that in the semigroup $(\mathbb{N}, +)$, the equation

$$n + x = m$$

is not solvable for arbitrary $m, n \in \mathbb{N}$. If $m \geq n$, then the unique solution is given by the difference $x = m - n$. If, on the other hand, we have $m < n$, then there is no solution in the set of natural numbers. This difficulty will now be overcome by extending the semigroup $(\mathbb{N}, +)$ to a group (G, \circ_G) , by which we mean that $\mathbb{N} \subseteq G$, and the restriction of the operation \circ_G to the subset \mathbb{N} coincides with the operation of addition $+$. Under these conditions, the

equation $n + x = m$ becomes transformed as an equation in G to $n \circ_G x = m$, which has the unique solution

$$x = n^{-1} \circ_G m.$$

Since the solution x in the case of $m < n$ cannot be a natural number, it must reside in $G \setminus \mathbb{N}$, the complement of \mathbb{N} in G .

We may thus inquire more generally into the circumstances under which it is possible to extend a semigroup (H, \circ_H) to a group (G, \circ_G) , namely a group G containing H such that the restriction of \circ_G to H coincides with the operation \circ_H . The following definition of *regular* semigroup is the key concept.

Definition 6.1. A semigroup (H, \circ_H) is said to be *regular* if for all elements $h, x, y \in H$, we have the *cancellation laws*

$$\begin{aligned} h \circ_H x = h \circ_H y &\implies x = y, \\ x \circ_H h = y \circ_H h &\implies x = y. \end{aligned}$$

Remark 6.2. (i) If the regular semigroup (H, \circ_H) is abelian, then we require only a single cancellation law in Definition 6.1.

(ii) A group (G, \circ_G) is itself a regular semigroup, since applying the inverse h^{-1} to $h \circ_G x = h \circ_G y$ ($h, x, y \in G$) from the left yields

$$h^{-1} \circ_G h \circ_G x = h^{-1} \circ_G h \circ_G y \iff x = y.$$

The other implication follows from applying the group operation with h^{-1} from the right.

Example 6.3. It is easy to show by mathematical induction that the semigroup $(\mathbb{N}, +)$ is regular. Because $(\mathbb{N}, +)$ is abelian, it suffices to prove the implication

$$h + x = h + y \implies x = y \quad (h, x, y \in \mathbb{N}). \quad (8)$$

To this end, fix $x, y \in \mathbb{N}$ and apply induction on h . For $h = 0$, the assertion is obviously correct, which establishes the basis of the induction. As induction hypothesis, we assume that the implication (8) is true for some $h \in \mathbb{N}$. We must then prove the implication

$$h^* + x = h^* + y \implies x = y$$

for the successor h^* of h . From the equation

$$(h + x)^* = h^* + x = h^* + y = (h + y)^*,$$

we obtain, on account of the injectivity of the successor mapping, that $h + x = h + y$, which yields $x = y$ at once by the induction hypothesis. Since $x, y \in \mathbb{N}$ were arbitrary, we have proved by induction the validity of the cancellation law in Definition 6.1 for all $h, x, y \in \mathbb{N}$.

Exercise 6.4.

- (a) Let A be a set with at least two elements. Show that neither of the two cancellation laws holds in the semigroup $(\text{map}(A), \circ)$.
- (b) Find other examples of semigroups that are not regular.

Theorem 6.5. *For every regular abelian semigroup (H, \circ_H) there exists a unique abelian group (G, \circ_G) satisfying the following two conditions:*

- (i) *H is a subset of G , and the restriction of \circ_G to H coincides with the operation \circ_H .*
- (ii) *If $(G', \circ_{G'})$ is another group satisfying property (i), then G is a subgroup of G' .*

Proof. We must prove both existence and uniqueness. We begin with a proof of uniqueness.

Uniqueness: Let (G_1, \circ_{G_1}) and (G_2, \circ_{G_2}) be groups satisfying properties (i) and (ii). By property (ii), we have in particular that $G_1 \leq G_2$, but conversely also that $G_2 \leq G_1$. That is, the two groups are identical. Therefore, the group in question is determined uniquely (up to isomorphism).

Existence: We begin by defining a relation \sim on the Cartesian product

$$H \times H = \{(a, b) \mid a, b \in H\}$$

(for simplicity of notation, we shall write \circ instead of \circ_H):

$$(a, b) \sim (c, d) \iff a \circ d = b \circ c \quad (a, b, c, d \in H).$$

We can easily show that this is an equivalence relation.

(a) *Reflexivity:* Since the semigroup (H, \circ) is abelian, it follows that $a \circ b = b \circ a$ for all $a, b \in H$. That is, $(a, b) \sim (a, b)$. Therefore, the relation \sim is reflexive.

(b) *Symmetry:* Let $(a, b), (c, d) \in H \times H$ be such that $(a, b) \sim (c, d)$, that is, $a \circ d = b \circ c$. Since (H, \circ) is abelian, we may conclude that $c \circ b = d \circ a$, which means precisely that $(c, d) \sim (a, b)$; that is, \sim is symmetric.

(c) *Transitivity:* Let $(a, b), (c, d), (e, f) \in H \times H$ be such that $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$. We have, therefore, the equalities

$$a \circ d = b \circ c, \quad c \circ f = d \circ e.$$

If we apply the group operation to the left-hand and right-hand sides of these two equations, we obtain, using the associativity and commutativity of the semigroup (H, \circ) , the following equivalent equalities:

$$\begin{aligned}
(a \circ d) \circ (c \circ f) &= (b \circ c) \circ (d \circ e), \\
a \circ d \circ c \circ f &= b \circ c \circ d \circ e, \\
(a \circ f) \circ (d \circ c) &= (b \circ e) \circ (d \circ c).
\end{aligned}$$

Since the semigroup (H, \circ) is also regular, we can cancel $(d \circ c)$ in the last equation (from the right), obtaining

$$a \circ f = b \circ e,$$

which implies $(a, b) \sim (e, f)$. The relation \sim is therefore also transitive.

We denote by $[a, b] \subseteq H \times H$ the equivalence class of the pair $(a, b) \in H \times H$, and by G the set of all such equivalence classes. For the sake of brevity, we write

$$G := (H \times H) / \sim.$$

Since the semigroup (H, \circ) is nonempty, so that it contains at least one element h , it follows that the set G is also nonempty, since it contains at least the equivalence class $[h, h]$. We now define an operation on the set G of equivalence classes, which for simplicity we shall denote by \bullet instead of \circ_G . If $[a, b], [a', b'] \in G$, then we define

$$[a, b] \bullet [a', b'] := [a \circ a', b \circ b'].$$

Since this definition apparently depends on the choice of representatives a, b and a', b' of the equivalence classes $[a, b]$ and $[a', b']$, we must prove that the operation \bullet is well defined by showing that it is, in fact, independent of this choice. To this end, let (c, d) and (c', d') be arbitrary representatives of $[a, b]$ and $[a', b']$. We must show that

$$[a \circ a', b \circ b'] = [c \circ c', d \circ d'] \iff (a \circ a', b \circ b') \sim (c \circ c', d \circ d').$$

Since we have $(c, d) \in [a, b]$ and $(c', d') \in [a', b']$, we must have

$$a \circ d = b \circ c \quad \text{and} \quad a' \circ d' = b' \circ c'.$$

By composing the left- and right-hand sides, we obtain, on the assumption of the commutativity of H ,

$$(a \circ d) \circ (a' \circ d') = (b \circ c) \circ (b' \circ c') \iff (a \circ a') \circ (d \circ d') = (b \circ b') \circ (c \circ c'),$$

and we have, therefore, as asserted,

$$(a \circ a', b \circ b') \sim (c \circ c', d \circ d').$$

In sum, we now have in (G, \bullet) a nonempty set with a well-defined operation. In the following four steps, we shall show that (G, \bullet) is an abelian group.

(1) We first show that \bullet is associative. But this can be shown easily from the definition of \bullet and the associativity of \circ with $[a, b], [a', b'], [a'', b''] \in G$:

$$\begin{aligned} ([a, b] \bullet [a', b']) \bullet [a'', b''] &= [a \circ a', b \circ b'] \bullet [a'', b''] \\ &= [(a \circ a') \circ a'', (b \circ b') \circ b''] = [a \circ (a' \circ a''), b \circ (b' \circ b'')] \\ &= [a, b] \bullet [a' \circ a'', b' \circ b''] = [a, b] \bullet ([a', b'] \bullet [a'', b'']). \end{aligned}$$

(2) The commutativity of \bullet follows equally easily from the commutativity of the operation \circ with $[a, b], [a', b'] \in G$:

$$[a, b] \bullet [a', b'] = [a \circ a', b \circ b'] = [a' \circ a, b' \circ b] = [a', b'] \bullet [a, b].$$

(3) We now show that G possesses an identity element. To this end, we choose an arbitrary element $h \in H$; we know that such an element exists, since H is nonempty. Then the equivalence class $[h, h]$ is our candidate for the identity element in G . Let $[a, b]$ be an arbitrary element of G . By the commutativity of \circ , we have

$$(h \circ a) \circ b = (h \circ b) \circ a \iff ((h \circ a), (h \circ b)) \sim (a, b).$$

Then from the commutativity of \bullet , we obtain

$$[a, b] \bullet [h, h] = [h, h] \bullet [a, b] = [h \circ a, h \circ b] = [a, b].$$

That is, $[h, h]$ is indeed the identity element in G .

(4) Finally, we must show that every element $[a, b] \in G$ has an inverse $[a, b]^{-1}$ in G . We assert that the desired inverse is given by $[b, a] \in G$. By the commutativity of \circ and \bullet , we see that

$$[a, b] \bullet [b, a] = [b, a] \bullet [a, b] = [b \circ a, a \circ b] = [a \circ b, a \circ b].$$

Now, since the equality $(a \circ b) \circ h = (a \circ b) \circ h$ is equivalent to $(a \circ b, a \circ b) \sim (h, h)$, we obtain the desired relation

$$[a, b] \bullet [b, a] = [b, a] \bullet [a, b] = [a \circ b, a \circ b] = [h, h].$$

To complete the proof, we must show that (G, \bullet) satisfies the two conditions (i), (ii) above, namely (i) that H is a subset of G and the restriction of \bullet to H coincides with the operation \circ , and (ii) that (G, \bullet) is minimal with respect to property (i).

To verify property (i), it suffices to find an injective mapping $f : H \rightarrow G$ satisfying

$$f(a \circ b) = f(a) \bullet f(b) \quad (a, b \in H). \quad (9)$$

By then identifying H with its image $f(H) \subseteq G$, we shall obtain, taking into account (9), the desired result. We define the mapping $f : H \rightarrow G$ by sending each element $a \in H$ to the element $[a \circ h, h] \in G$ (the element h was cho-

sen when we defined the identity element $[h, h]$ of G). We now show that f is injective. Let $a, b \in H$ be such that

$$f(a) = f(b) \iff [a \circ h, h] = [b \circ h, h] \iff (a \circ h, h) \sim (b \circ h, h).$$

Given the commutativity and regularity of (H, \circ) , we see that this is equivalent to

$$(a \circ h) \circ h = h \circ (b \circ h) \iff a \circ h^2 = b \circ h^2 \iff a = b,$$

from which the injectivity of f follows.

To prove (9), we choose two arbitrary elements $a, b \in H$ and calculate, taking into account the associativity and commutativity of \circ ,

$$\begin{aligned} f(a \circ b) &= [(a \circ b) \circ h, h] = [a \circ b \circ h, h] = [a \circ b \circ h \circ h, h \circ h] \\ &= [(a \circ h) \circ (b \circ h), h \circ h] = [a \circ h, h] \bullet [b \circ h, h] = f(a) \bullet f(b). \end{aligned}$$

We have thereby demonstrated the structure-preserving property (9) of f , showing that (G, \bullet) is an abelian group satisfying property (i).

To complete the proof, we show that the group (G, \bullet) that we have constructed is minimal. To this end, we show that the group (G, \bullet) cannot be made any smaller. By identifying, as mentioned above, the semigroup (H, \circ) with its image in (G, \bullet) under f , we see that by construction, G must contain all elements of the form $[a \circ h, h]$ for $a \in H$. Since (G, \bullet) is a group, it must contain for each such $[a \circ h, h]$ the inverse $[h, a \circ h]$ in G ; that is, G also contains all elements of the form $[h, b \circ h]$ with $b \in H$. Because G is closed under the operation \bullet , it must also contain all elements of the form

$$[a \circ h, h] \bullet [h, b \circ h] = [a, b] \quad (a, b \in H).$$

But this shows that one cannot omit a single equivalence class from G , showing that (G, \bullet) is minimal. \square

Exercise 6.6.

- (a) Show that the odd natural numbers under multiplication form a regular abelian monoid.
- (b) Carry out the construction for this monoid described in Theorem 6.5.

7. The Integers

We would like to investigate more closely the abelian group (G, \circ_G) constructed in Theorem 6.5 using the example of the regular abelian semigroup $(H, \circ_H) = (\mathbb{N}, +)$. In doing so, we shall introduce the set of *integers*.

We begin by noting that the equivalence relation \sim defined on the Cartesian product $\mathbb{N} \times \mathbb{N}$ now assumes the form

$$(a, b) \sim (c, d) \iff a + d = b + c \quad (a, b, c, d \in \mathbb{N}).$$

The abelian group (G, \circ_G) is given, according to the proof of Theorem 6.5, by the set of all equivalence classes $[a, b]$ associated with pairs $(a, b) \in \mathbb{N} \times \mathbb{N}$ and is equipped with the operation

$$[a, b] \circ_G [a', b'] = [a + a', b + b'] \quad ([a, b], [a', b'] \in G);$$

the identity element in (G, \circ_G) is given by the element $[0, 0]$, where 0 denotes the natural number zero. Since we are dealing here with an additive structure, we shall write the inverse $[a, b]^{-1}$ in the form $-[a, b]$.

The definition of the equivalence relation \sim shows in this special case that every equivalence class can be expressed in the form

$$[a, b] = \begin{cases} [a - b, 0], & \text{if } a \geq b, \\ [0, b - a], & \text{if } b > a. \end{cases}$$

We see, then, that the underlying set G of the group (G, \circ_G) is given by the union

$$G = \{[n, 0] \mid n \in \mathbb{N}\} \cup \{[0, n] \mid n \in \mathbb{N}\},$$

where the intersection $\{[n, 0] \mid n \in \mathbb{N}\} \cap \{[0, n] \mid n \in \mathbb{N}\}$ consists solely of the identity element $[0, 0]$. We see from the proof of Theorem 6.5 that the set of natural numbers \mathbb{N} is in bijection with the set $\{[n, 0] \mid n \in \mathbb{N}\}$. This bijection is induced by the assignment $n \mapsto [n, 0]$. By identifying the set of natural numbers \mathbb{N} with the set $\{[n, 0] \mid n \in \mathbb{N}\}$, that is, we set $n = [n, 0]$, we may henceforth view \mathbb{N} as a subset of G .

Definition 7.1. For a nonzero natural number n , we now set

$$-n := [0, n].$$

Taking into account the identification of \mathbb{N} with $\{[n, 0] \mid n \in \mathbb{N}\}$ and using the previous definition, we can realize G in the form

$$G = \{0, 1, 2, 3, \dots\} \cup \{-1, -2, -3, \dots\}.$$

Definition 7.2. We shall hereinafter denote the group (G, \circ_G) by $(\mathbb{Z}, +)$ and call it the *(additive) group of integers*. As a set, we may represent \mathbb{Z} in the form

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

We call the numbers $1, 2, 3, \dots$ *positive integers*, the numbers $-1, -2, -3, \dots$ *negative integers*. Finally, for the integer given by the equivalence class $[a, b]$, we introduce the usual notation

$$a - b := [a, b]$$

and call it the *difference* of the natural numbers a and b .

Remark 7.3. (i) Definition 7.2, which defines the difference of two natural numbers, is unrestricted and therefore generalizes the notion of difference given in Definition 1.24 of Chapter I. Moreover, the general notion of difference in Definition 7.2 is compatible with the notion of difference in Definition 1.24 of Chapter I: if $a, b \in \mathbb{N}$ with $a \geq b$, then by Definition 7.2, we have $a - b = [a, b]$. Using Definition 1.24 of Chapter I, this can be transformed into $a - b = [a - b, 0]$; the identification of \mathbb{N} with $\{[n, 0] \mid n \in \mathbb{N}\}$ now shows the asserted compatibility.

(ii) Since we denote the inverse $[a, b] = a - b$ by $-[a, b] = -(a - b)$, which is in turn given by $[b, a] = b - a$, we obtain

$$-(a - b) = b - a.$$

If we set $a = 0$, we obtain in particular the formula $-(-b) = b$ ($b \in \mathbb{N}$).

(iii) Using (ii), we now obtain in general the *difference of two integers* $a - b = [a, b]$ and $a' - b' = [a', b']$ in the form

$$(a - b) - (a' - b') := (a - b) + (-(a' - b')) = (a - b) + (b' - a').$$

(iv) One should keep in mind in considering the difference $a - b$ that there is always an equivalence class lurking in the background; for example,

$$-2 = 1 - 3 = 2 - 4 = 3 - 5 = \dots;$$

that is, the pairs of natural numbers $(1, 3), (2, 4), (3, 5), \dots$ are all representatives of the integer -2 and of the equivalence class $[0, 2]$.

Definition 7.4. We extend the relation \leq on the set \mathbb{N} of natural numbers given in Definition 1.15 of Chapter I to the set \mathbb{Z} of integers by declaring that every negative integer is strictly less than every natural number and that for two negative integers $-m, -n$ ($m, n \in \mathbb{N}; m, n \neq 0$), we set

$$\begin{aligned} -m < -n & \text{ if } m > n, \\ -m \leq -n & \text{ if } m \geq n. \end{aligned}$$

We extend the relations $>$ and \geq to the set \mathbb{Z} of integers analogously.

In analogy to Remark 1.16 of Chapter I, we have the following.

Remark 7.5. With the relation $<$, the set of integers \mathbb{Z} is an *ordered set*; that is, the following conditions are satisfied:

- (i) For every two elements $m, n \in \mathbb{Z}$, we have $m < n$ or $n < m$ or $m = n$.
- (ii) The three relations $m < n, n < m, m = n$ are mutually exclusive.
- (iii) If $m < n$ and $n < p$, then $m < p$.

Analogous properties hold for $>$.

Exercise 7.6. Generalize the addition and multiplication rules for the natural numbers in Remark 1.19 of Chapter I to the set of integers.

Definition 7.7. Let $n \in \mathbb{Z}$ be an integer. We then set

$$|n| := \begin{cases} n, & \text{if } n \geq 0, \\ -n, & \text{if } n < 0. \end{cases}$$

We call the natural number $|n|$ the *absolute value* of the integer n .

Example 7.8. The set of integers $(\mathbb{Z}, +)$ with the operation of addition that we have constructed gives us an additional example of an abelian group. If $n \in \mathbb{N}$ is a nonzero natural number, then the set

$$n\mathbb{Z} = \{\dots, -3n, -2n, -n, 0, n, 2n, 3n, \dots\}$$

of all integral multiples of n forms a subgroup $(n\mathbb{Z}, +)$ of $(\mathbb{Z}, +)$. Since $(\mathbb{Z}, +)$ is an abelian group, the subgroup $(n\mathbb{Z}, +)$ will automatically be a normal subgroup of $(\mathbb{Z}, +)$, and we can consider the quotient group $(\mathbb{Z}/n\mathbb{Z}, \bullet)$.

Furthermore, we may easily verify that the assignment $a \mapsto R_n(a)$ ($a \in \mathbb{Z}$) induces a group homomorphism

$$f: (\mathbb{Z}, +) \longrightarrow (\mathcal{R}_n, \oplus).$$

This group homomorphism f is obviously surjective, and its kernel is

$$\ker(f) = n\mathbb{Z}.$$

The corollary to the homomorphism theorem for groups yields for us the group isomorphism

$$(\mathbb{Z}/n\mathbb{Z}, \bullet) \cong (\mathcal{R}_n, \oplus);$$

here the coset $a + n\mathbb{Z} \in \mathbb{Z}/n\mathbb{Z}$ is mapped to the element $R_n(a) \in \mathcal{R}_n$. This example demonstrates nicely how the complicated structure of the quotient group $(\mathbb{Z}/n\mathbb{Z}, \bullet)$ that we have been gradually developing can be identified with the simple n -element set \mathcal{R}_n , on which we may perform “addition” by taking remainders.

Exercise 7.9. Verify the assertions of this example in detail.

Remark 7.10. Theorem 6.5 applied to the regular abelian monoid $(\mathbb{N} \setminus \{0\}, \cdot)$ yields the multiplicative group of *fractions* (\mathbb{B}, \cdot) . We shall not discuss the group (\mathbb{B}, \cdot) further, since in Section 6 of Chapter III, we shall rediscover this group as the multiplicative group of positive rational numbers.

B. RSA Encryption: An Application of Number Theory

In this final section, we shall discuss the ideas behind RSA encryption as an interesting and current application of the properties of the integers.

B.1 Cryptography

The purpose of cryptography (from the Greek *kryptos*, hidden, and *graphos*, writing) is to maintain secrecy in communication so that unauthorized agents are unable to read or alter a message while it is being transmitted from sender to receiver. The basic principle is simple. The unencrypted message, or plaintext, is transformed with the help of a key into a ciphertext that is no longer comprehensible. Only someone in possession of the key can decrypt the ciphertext back into the original plaintext, thereby making the message understandable.

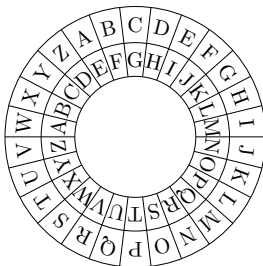
The history of cryptography goes back at least to the second century B.C., encrypted texts having been found as inscriptions on tombstones from that period. We are not, however, going to delve into the history of the subject, for which we refer the reader to the relevant literature, some of which is of a popular nature (see, for example, [2, 7]). We shall instead touch on some of the basic ideas behind encryption algorithms.

In symmetric encryption algorithms, the keys for encryption and decryption are essentially the same. For example, the key to such an algorithm might consist in replacing each plaintext letter by a uniquely determined ciphertext letter.

A well-known example is the *Caesar cipher*, whereby the letters of the alphabet in the top row are displaced cyclically by a certain number of places:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

If, for example, that number of places is four, then the plaintext “HAIL CAESAR” would be encrypted as “LEMP GEIWEV,” as can be seen in the following graphic, in which the plaintext letters appear in the outside ring, and their corresponding ciphertext letters can then be read off on the inside ring. This variant of the Caesar cipher is quite simple, but the price of that simplicity is that it is a very insecure encryption technique, since for an alphabet of twenty-six letters, there are only twenty-five different possible keys, so that without even the use of frequency analysis of the letters, the ciphertext could be decrypted after at most the twenty-fifth attempt.



The security of this method rested primarily, in the period when it was used, on the fact that the method of encryption was kept secret.

In modern cryptography, in contrast, a fundamental principle, called *Kerckhoffs's principle*, after the Dutch linguist and cryptographer Auguste Kerckhoffs, states that the security of an encryption algorithm should depend only on the security of the key and not on the secrecy of the algorithm.

A polyalphabetic modification of the Caesar cipher is the *Vigenère cipher*, named for Blaise de Vigenère, which uses an additional keyword to determine the number of offset letters in the Caesar cipher. A popular convention for this method is that the letter A in the keyword represents no offset; the letter B, an offset of 1; the letter C, an offset of 2, and so on. If, for example, the supplementary keyword is "FANATIC," then the alphabet for the first letter to be encrypted is offset by 5; for the second letter, by 0; for the third, by 13, and so on, 0, 19, 8, and 2. If the plaintext message is again "HAIL CAESAR," then the ciphertext will read "MAVL VIGXAR." While much more secure than the Caesar cipher, this encryption method is truly secure only if the supplementary keyword is the same length as the plaintext, which in general makes for a great deal of overhead.

Another variety of a polyalphabetic encryption algorithm is the basis of the famous *Enigma* code, which used a sort of electromechanical typewriter, making rapid encryption and decryption possible. The plaintext would be input by keyboard. Then the letters of the plaintext were passed to three rotors, a reflector, and again three rotors, with the encrypted ciphertext finally displayed on a lamp board. The Enigma code was used by the Germans in the Second World War, and was considered, incorrectly, to be unbreakable.

With the critical assistance of the mathematician Alan Turing, the British were able to crack encrypted German radio messages beginning in about 1940. An extensive description of Enigma, including its weaknesses and possible improvements, can be found, for example, in [2] and in [7]. There are also several enjoyable films on this topic, including the 2014 biopic "The Imitation Game."

The reason that mathematics plays such an important role in cryptography is that there are many ways of encoding the information to be encrypted in the form of a number or sequence of numbers. For example, one can encode the alphabet using ASCII encoding as follows:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90

Thus the text “HAIL CAESAR” corresponds in ASCII to the sequence

72, 65, 73, 76, 67, 65, 69, 83, 65, 82,

or simply to the number 72657376676569836582. Once the plaintext has been written in the form of a number, encryption becomes a mathematical function whose uniquely defined inverse is the function for decryption.

If in the ASCII coding above, we replace each number by its remainder on division by 26, we obtain the following encoding substitutions:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
13	14	15	16	17	18	19	20	21	22	23	24	25	0	1	2	3	4	5	6	7	8	9	10	11	12

and now “HAIL CAESAR” will be encoded as 20,13,21,24,15,13,17,5,13,4. If we now employ a Caesar cipher with offset 4, the encryption function described above amounts to adding 4 to the corresponding number and subtracting 26 if the resulting number is greater than 25.

The plaintext 20,13,21,24,15,13,17,5,13,4 will therefore be encoded as 24,17,25,2,19,17,9,17,8. Decryption involves simply applying the inverse function, that is, subtracting the number 4 and then adding 26 if the resulting number is negative. It is mathematically more elegant to describe the encryption function and its inverse in terms of congruences (modulo 26), which we shall learn about in Section B.2.

In the case of symmetric encryption procedures, it is easy in general to obtain the decryption function from knowledge of the encryption function, whence the name “symmetric.” There is, however, a fundamental problem regarding the security of symmetric encryption algorithms, namely to send the decryption function to the recipient of the ciphertext over a secure channel. If an enemy can tap into the transmission and obtain the decryption function, then the security of the ciphertext will have been compromised.

In 1976, Whitfield Diffie und Martin Hellman proposed in the article [4] that the problem of security might be solved by using two different keys, a *public key* for encryption, which is available to everyone, and a *private key* for decryption, which must remain secret, known only by the recipient of the ciphertext. This idea proved decisive for the transition from classical cryptography to the modern concept of *public key cryptography*.

To realize this idea mathematically, the encryption function must have the property that an adversary would find it impossible to compute the inverse

function from knowledge of the encryption function without additional information; even if the attacker could theoretically compute the inverse function, it would take so long that in practice, it could not be done. For the recipient, however, who is in possession of the private key, computing the inverse function is easy. Moreover, the encryption function should have the additional property that it is easy to convert plaintext to ciphertext, say in polynomial time. Such a function is called a *one-way trapdoor function*.

The question of the existence of such a one-way trapdoor function remained long unresolved until three computer scientists, Ronald Rivest, Adi Shamir, and Leonard Adleman, attempted to show that no such function could exist. But instead of doing so, they in fact discovered such a function. In 1977, they produced an encryption algorithm known today by the initials of their three surnames, the *RSA algorithm*, the first published asymmetric encryption algorithm; see [6]. Independently, similar ideas were developed four years earlier by mathematicians in the British secret service, among them Clifford Cocks and James Ellis. Their work, however, was not published.

Today, the RSA algorithm is a widely used asymmetric procedure, with applications in telephony, electronic banking, credit-card transactions, and in the Internet, for example in email encryption and transmission protocols such as TLS and SSH.

In the following sections, we provide a glimpse into how the RSA algorithm works along with the necessary elementary number theory. In particular, we discuss congruence arithmetic, for which we shall need the theory of divisibility and the Euclidean algorithm for the ring $(\mathbb{Z}, +, \cdot)$ from Chapter III.

B.2 Congruence Arithmetic

In this section, we introduce the notion of *congruence arithmetic*. We begin by defining a relation on the set \mathbb{Z} of integers.

Definition B.1. Let $m \in \mathbb{N}$ and $m > 0$. For $a, b \in \mathbb{Z}$, we define

$$a \equiv b \pmod{m} \iff m \mid (b - a)$$

and say that a is *congruent to b modulo m* . The relation \equiv is called *congruence modulo m* .

Remark B.2. If $a \equiv c \pmod{m}$ and $b \equiv d \pmod{m}$, then we have the following two rules:

- (i) $a + b \equiv c + d \pmod{m}$,
- (ii) $a \cdot b \equiv c \cdot d \pmod{m}$.

Rule (ii) shows in particular that if $a \equiv c \pmod{m}$, then for every $n \in \mathbb{N}$, we have the congruence

$$a^n \equiv c^n \pmod{m}.$$

Example B.3. Let $m = 22$. Then, for example, we have $23 \equiv 1 \pmod{22}$, $47 \equiv 3 \pmod{22}$, and $87 \equiv 21 \pmod{22}$. Using the above rules of calculation, we obtain $23 + 47 \equiv 1 + 3 \equiv 4 \pmod{22}$ and $47 \cdot 87 \equiv 3 \cdot 21 \equiv 19 \pmod{22}$, as well as $47^{17} \equiv 3^{17} \equiv 129140163 \equiv 9 \pmod{22}$.

Remark B.4. For larger numbers, one can make use of freely available mathematical software such as SAGE (www.sagemath.org). You can calculate a modulo m with the command `mod(a,m)` and the exponentiation a to the power n modulo m with `power_mod(a,n,m)`.

The command `power_mod(a,n,m)` is implemented in such a way as to minimize the number of multiplications (*square-and-multiply algorithm*), thereby computing exponentials in minimal time. This is of practical importance, since modern cryptosystems often require the rapid calculation of powers modulo m . You can see the difference by performing a test calculation on large numbers using `power_mod(a,n,m)` and `mod(a^n,m)`.

Remark B.5. For $a, b \in \mathbb{Z}$, we clearly have the equivalence

$$a \equiv b \pmod{m} \iff R_m(a) = R_m(b).$$

Thus a is congruent to b modulo m if and only if a and b have the same remainder on division by m . The calculational rules given above show that calculation with congruences is easier than with remainders.

It is easy to show that the relation \equiv is an equivalence relation on the set \mathbb{Z} of integers. The equivalence class of $a \in \mathbb{Z}$ is called the *residue class of a modulo m* and is denoted by \bar{a} or $a \pmod{m}$. The residue classes modulo m are given by the set

$$\{\bar{0}, \bar{1}, \dots, \overline{m-1}\},$$

which stands in natural bijection with the set \mathcal{R}_m of remainders on division by m as shown in Example 7.8.

Theorem B.6. For a given integer a , the congruence

$$a \cdot x \equiv 1 \pmod{m} \tag{10}$$

has a solution for $x \in \mathbb{Z}$ if and only if $(a, m) = 1$. If that is the case and if $x \in \mathbb{Z}$ is a solution to the congruence (10), then that congruence can be solved precisely for all integers $x' \in \mathbb{Z}$ such that $x' \equiv x \pmod{m}$.

Proof. The solvability of the congruence (10) is equivalent to that of the equation

$$a \cdot x + m \cdot y = 1$$

for $x \in \mathbb{Z}$ (and some $y \in \mathbb{Z}$). If d is a common divisor of a and m , then we must have the divisibility relationship $d \mid 1$, which proves the equality $(a, m) = 1$.

We now show that this condition is also sufficient for solving the congruence (10). Since we have $(a, m) = 1$, there exist, by the extended Euclidean algorithm (see Remark 7.36 of Chapter III), $x, y \in \mathbb{Z}$ such that

$$a \cdot x + m \cdot y = 1.$$

But this means that for x , we have the congruence

$$a \cdot x \equiv 1 \pmod{m},$$

and x is therefore a solution of the congruence (10). If we now have a further solution $x' \in \mathbb{Z}$ of the congruence (10), then we have the equivalence

$$a \cdot x \equiv 1 \equiv a \cdot x' \pmod{m} \iff a(x - x') \equiv 0 \pmod{m}.$$

Because of the relative primality of a and m , we must have that m divides the difference $x - x'$; that is, we have

$$x' \equiv x \pmod{m}.$$

This proves that the congruence (10) is solved precisely by all numbers $x' \in \mathbb{Z}$ such that $x' \equiv x \pmod{m}$. \square

Example B.7. Let $m = 88464$ and $a = 43$. Since 43 is prime and m is not a multiple of 43, we have $(43, 88464) = 1$. Using the Euclidean algorithm (see Theorem 7.35 of Chapter III), we obtain, by repeated division with remainder,

$$\begin{aligned} 88464 &= 2057 \cdot 43 + 13, \\ 43 &= 3 \cdot 13 + 4, \\ 13 &= 3 \cdot 4 + 1, \\ 4 &= 4 \cdot 1 + 0, \end{aligned}$$

which verifies that $(43, 88464) = 1$. If we perform this calculation in reverse, we obtain

$$\begin{aligned} 1 &= 13 - 3 \cdot 4, \\ 1 &= 13 - 3 \cdot (43 - 3 \cdot 13) = 10 \cdot 13 - 3 \cdot 43, \\ 1 &= 10 \cdot (88464 - 2057 \cdot 43) - 3 \cdot 43 = 10 \cdot 88464 - 20573 \cdot 43. \end{aligned}$$

Thus the congruence $43 \cdot x \equiv 1 \pmod{88464}$ is solved by

$$x \equiv -20573 \pmod{88464}.$$

Remark B.8. For larger numbers, this calculation can be carried out with the SAGE command `xgcd(a,m)`. For example, we have

$$\text{xgcd}(43, 88464) = (1, -20573, 10),$$

which means that $(43, 88464) = 1$ and that we have the equality

$$43 \cdot (-20573) + 88464 \cdot 10 = 1.$$

Finally, if we calculate

$$\text{mod}(-20573, 88464) = 67891,$$

we obtain a solution x such that $0 < x < 88464$.

Remark B.9. A solution x of the congruence $a \cdot x \equiv 1 \pmod{m}$ with $0 < x < m$ can also be obtained with the SAGE command `a.inverse_mod(m)`. For example, for $a = 43$ and $m = 88464$, we obtain the result

$$43.\text{inverse_mod}(88464) = 67891.$$

B.3 Theorems of Fermat and Euler

The following theorem is due to the French mathematician Pierre de Fermat.

Theorem B.10 (Fermat's little theorem). *Let p be a prime number. Then for all integers a , we have the congruence*

$$a^p \equiv a \pmod{p}.$$

Proof. If a is a multiple of p , then we have $a \equiv 0 \pmod{p}$ and therefore also $a^p \equiv 0 \pmod{p}$, which proves the asserted congruence in this case.

If, on the other hand, a is not a multiple of p , then a is relatively prime to p . We now consider the product $a \cdot j$ with $j \in \{1, \dots, p-1\}$. Since both a and j are relatively prime to p , division by p with remainder shows that there exists $j' \in \{1, \dots, p-1\}$ such that

$$a \cdot j \equiv j' \pmod{p}.$$

The assignment $j \mapsto j'$ clearly induces a mapping of the set $\{1, \dots, p-1\}$ to itself. This map is injective, since the equivalence

$$a \cdot j_1 \equiv j' \pmod{p} \iff a \cdot j_2 \equiv j' \pmod{p} \iff p \mid a(j_1 - j_2)$$

and the relative primality of a and p immediately imply $j_1 = j_2$. Because the set $\{1, \dots, p-1\}$ is finite, injectivity implies surjectivity, and so the mapping under consideration is bijective. Taking products yields the congruence

$$(a \cdot 1) \cdots (a \cdot (p-1)) \equiv 1 \cdots (p-1) \pmod{p},$$

that is,

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod{p} \iff (a^{p-1} - 1)(p-1)! \equiv 0 \pmod{p}.$$

Since $(p-1)!$ is relatively prime to p , it follows from Euclid's lemma (Lemma 3.3 of Chapter I) that $p \mid (a^{p-1} - 1)$, which is equivalent to

$$a^{p-1} \equiv 1 \pmod{p}.$$

Multiplying this congruence by a establishes the statement of the theorem in this second case. \square

Remark B.11. Fermat's little theorem gives us, in particular, a simple primality test. If, for example, $a = 2$ and $m = 15$, then we can calculate $a^m \pmod{m}$, obtaining

$$2^{15} \equiv 2^5 \cdot 2^5 \cdot 2^5 \equiv 2^3 \equiv 8 \pmod{15}.$$

Since 8 is not congruent to 2 modulo 15, 15 cannot be prime.

The Swiss mathematician Leonhard Euler generalized Fermat's little theorem to a modulus m that is the product of two distinct primes.

Theorem B.12 (Euler's theorem). *Let p and q be two distinct prime numbers, and set $m = p \cdot q$. Then for every integer a , we have the congruence*

$$a^{(p-1)(q-1)+1} \equiv a \pmod{m}.$$

Proof. We distinguish four cases.

(i) The integer a is a multiple of both p and q . In this case, we have, for a suitable choice of $b \in \mathbb{Z}$, the equality $a = b \cdot p \cdot q$. This yields $a \equiv 0 \pmod{p \cdot q}$, and by the rules for calculating with congruences,

$$a^{(p-1)(q-1)+1} \equiv 0 \pmod{p \cdot q},$$

which proves the asserted congruence.

(ii) The integer a is a multiple of p , but not of q . In this case, we have $a \equiv 0 \pmod{p}$, and therefore also $a^{(p-1)(q-1)+1} \equiv 0 \pmod{p}$, which leads to the congruence

$$a^{(p-1)(q-1)+1} \equiv a \pmod{p}. \tag{11}$$

But since a is not a multiple of q , the integer $b := a^{p-1}$ is relatively prime to q , and from the second part of the proof of Fermat's little theorem, we obtain the congruence $b^{q-1} \equiv 1 \pmod{q}$, that is,

$$a^{(p-1)(q-1)} \equiv 1 \pmod{q}.$$

On multiplication by a , we obtain the congruence

$$a^{(p-1)(q-1)+1} \equiv a \pmod{q}. \quad (12)$$

Since the prime numbers p and q are distinct, the two congruences (11) and (12) together yield the asserted congruence

$$a^{(p-1)(q-1)+1} \equiv a \pmod{p \cdot q}.$$

(iii) (The integer a is a multiple of q , but not of p . This case can be reduced to the previous case by interchanging the roles of p and q .)

(iv) The integer a is a multiple of neither p nor q . As in case (ii), one proves the congruence

$$a^{(p-1)(q-1)+1} \equiv a \pmod{q}. \quad (13)$$

Analogously, one proves the further congruence

$$a^{(p-1)(q-1)+1} \equiv a \pmod{p}. \quad (14)$$

Since the two primes p and q are distinct, we obtain from the congruences (13) and (14) the congruence

$$a^{(p-1)(q-1)+1} \equiv a \pmod{p \cdot q}.$$

This completes the proof. \square

Remark B.13. The theorems of Fermat and Euler presented here are special cases of a more general result, which derives ultimately from Lagrange's theorem (Theorem 4.14). Namely, if (G, \circ) is a finite group with identity element e , then every element $g \in G$ satisfies the relation $g^{|G|} = e$.

We now apply this result. We begin with the set

$$P(m) := \{\bar{a} \mid a \in \{0, \dots, m-1\}, (a, m) = 1\}.$$

It is now easy to see that the set $P(m)$ with respect to congruence multiplication is a group with identity element $\bar{1}$, whose order is usually denoted by $\varphi(m)$, called *Euler's φ -function*. By the previous result, we have for all $\bar{a} \in P(m)$, the relation

$$\bar{a}^{\varphi(m)} = \bar{1}.$$

We have, therefore, for all $a \in \mathbb{Z}$ with $(a, m) = 1$, the congruence

$$a^{\varphi(m)} \equiv 1 \pmod{m},$$

from which on multiplication by a , we obtain the congruence $a^{\varphi(m)+1} \equiv a \pmod{m}$.

We obtain the connection to the theorems of Euler and Fermat by verifying the formulas

$$\varphi(p) = p - 1 \quad \text{and} \quad \varphi(p \cdot q) = (p - 1)(q - 1),$$

for distinct primes p and q .

B.4 The RSA Cryptosystem

In this section, we shall learn about the ideas behind the RSA cryptosystem. For information on important and interesting questions, in particular security, including the choice of suitable prime numbers and the private key, as well as possible attacks against RSA, we refer the reader to the enormous literature on the subject. We note here that the examples presented in this section serve a pedagogical purpose and are not of any practical utility.

To send an encrypted message using the RSA algorithm, sender Alice and recipient Bob proceed as follows:

1. Before a message can be encoded and sent, Bob does the following: He chooses two distinct “large” prime numbers p and q , of approximately three hundred digits, which must be kept secret. Bob then computes the products

$$\begin{aligned} m &= p \cdot q, \\ n &= (p - 1) \cdot (q - 1); \end{aligned}$$

note that with the Euler φ -function, we have $n = \varphi(m)$. Now Bob chooses a natural number k that is relatively prime to n . The numbers m and k comprise the *public key*, and recipient Bob sends this information to sender Alice. Bob keeps the numbers p , q , and n private.

2. Alice now begins by transforming the message she wishes to send into a number a , using, for example, the ASCII code described earlier, with the properties

$$(a, m) = 1 \quad \text{and} \quad 0 < a < m.$$

If it turns out that $a \geq m$, the message can be split up into several blocks of suitable size so that in each of them, one has $a < m$. Alice then encrypts her message a by calculating the uniquely determined number b such that

$$b \equiv a^k \pmod{m} \quad \text{and} \quad 0 < b < m.$$

Now Alice sends Bob the encrypted message b over a channel that does not have to be secure.

3. To decrypt the ciphertext b , Bob now determines the uniquely determined (by Theorem B.6) integer x such that

$$k \cdot x \equiv 1 \pmod{n} \quad \text{and} \quad 0 < x < n.$$

Using the *private key* x , Bob computes the uniquely determined integer c such that

$$c \equiv b^x \pmod{m} \quad \text{and} \quad 0 < c < m.$$

The ciphertext has now been decoded, since $c = a$, as the following theorem establishes.

Theorem B.14. *With the above notation and assumptions, we have the equality*

$$a = c.$$

Proof. Since $0 < a, c < m$, we shall have the asserted equality $a = c$ once the existence of the congruence $c \equiv a \pmod{m}$ has been validated. We can see this as follows: We have the congruences $c \equiv b^x \pmod{m}$ and $b \equiv a^k \pmod{m}$, and therefore,

$$c \equiv (a^k)^x \equiv a^{k \cdot x} \pmod{m}. \quad (15)$$

Here the integer x is uniquely determined by the conditions $k \cdot x \equiv 1 \pmod{n}$ and $0 < x < n$; that is, there exists, in particular, a uniquely determined integer y with

$$k \cdot x = 1 + n \cdot y.$$

From (15), we obtain the congruences

$$c \equiv a^{k \cdot x} \equiv a^{1+n \cdot y} \equiv a \cdot (a^n)^y \pmod{m}.$$

Since we now have $(a, m) = 1$, the proof of Euler's theorem shows that we have the congruence $a^n \equiv 1 \pmod{m}$, which yields, finally, the congruence

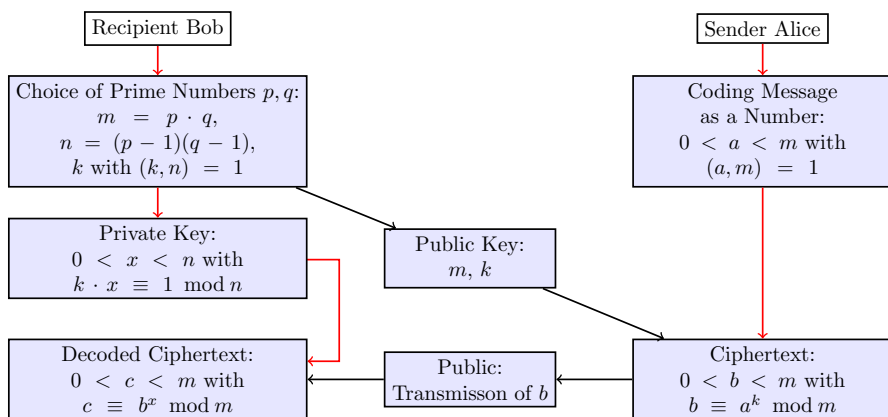
$$c \equiv a \cdot 1^y \equiv a \pmod{m},$$

which is what was to be proved. □

Remark B.15. It is possible in the second step of the RSA algorithm to do without the requirement $(a, m) = 1$ in creating the ciphertext. The correctness of the RSA algorithm was proved essentially in the previous proof.

Remark B.16. The encryption function used in the RSA algorithm is, in fact, a one-way trapdoor function (see Section B.1). Moreover, the process of encryption, which consists essentially in calculating $a^k \bmod m$, is simple. The same holds for calculating $b^x \bmod m$, provided, of course, that one knows the trapdoor, that is, the private key x . One can also easily compute the key x if like the recipient Bob, one knows the integer $\varphi(m) = n = (p - 1)(q - 1)$.

On the other hand, if one knows the public information m, k , and b , one could calculate x if one knew the prime decomposition of m , that is, the prime numbers p and q . If m is not particularly large, then one could figure out those prime factors, for example with the SAGE command `factor(m)`. But if m is very large, then deducing the prime decomposition is for all practical purposes impossible using currently known algorithms. Furthermore, if recipient Bob chooses, as is generally done, prime numbers of the same bit length (*balanced RSA algorithm*), then it is known that determining the private key x with only knowledge of the public information m, k , and b is of the same level of difficulty as the factorization of m . Thus the security of the RSA cryptosystem rests on the difficulty of the *factorization problem*.



The above diagram provides an overview of the RSA algorithm. The red arrows represent secure communication channels.

Example B.17. To aid in understanding the algorithm, we present here an example using two small prime numbers.

1. Recipient Bob chooses the prime numbers $p = 229$ and $q = 389$. He then calculates

$$m = p \cdot q = 229 \cdot 389 = 89081,$$

$$n = (p - 1) \cdot (q - 1) = 228 \cdot 388 = 88464.$$

Now Bob chooses, say, $k = 43$. Since 43 is a prime number and n is not a multiple of 43, we have that k is relatively prime to n , as desired. Bob now publishes the numbers

$$m = 89081 \quad \text{and} \quad k = 43.$$

2. Sender Alice transcribes the message “PI” using the ASCII encoding to obtain

$$a = 8073.$$

She then calculates the integer b with $b \equiv 8073^{43} \pmod{89081}$ and $0 < b < 89081$, and transmits the encrypted message

$$b = 30783.$$

3. To decode the ciphertext b , Bob computes the private key x such that $k \cdot x \equiv 1 \pmod{n}$ and $0 < x < n$ as in Example B.7. He thereby obtains the solution

$$x = 67891.$$

Bob can now decrypt the ciphertext $b = 30783$ by calculating the uniquely determined integer c such that $c \equiv 30783^{67891} \pmod{89081}$ and $0 < c < 89081$. He obtains

$$c = 8073,$$

which is the ASCII code for the message “PI”.

Example B.18. To close, we give a more realistic example with two 100-digit prime numbers.

1. Recipient Bob chooses the prime numbers

$$p = 20747222467734852078216952221076085874809964747211$$

$$17292752992589912196684750549658310084416732550077,$$

$$q = 72126101472954749095445237850434924099693821481867$$

$$65460082500085393519556525921455588705423020751421.$$

With the SAGE commands `is_prime(p)` and `is_prime(q)`, Bob can check whether p and q are in fact prime. The function returns `True` if its argument is prime. Bob then calculates the numbers m and n , obtaining

$$\begin{aligned}
m &= p \cdot q \\
&= 14964162729898105788684569421835754781481603923778 \\
&\quad 96104167832218033314436822709860751513251318961222 \\
&\quad 52290737219239160591728298144292465045647829035182 \\
&\quad 95622360979392187621542015444916226124162051409417
\end{aligned}$$

and

$$\begin{aligned}
n &= (p - 1) \cdot (q - 1) \\
&= 14964162729898105788684569421835754781481603923778 \\
&\quad 96104167832218033314436822709860751513251318961221 \\
&\quad 59417413278549559418066108072781455071144042806104 \\
&\quad 12869525486716881905300738973802327334322298107920.
\end{aligned}$$

At this point, the reader may wish to test the SAGE command `factor(m)`. Now Bob chooses again, for example, $k = 43$, since k is relatively prime to n , as desired. One can verify this with the SAGE command `gcd(n,k)`, which returns the greatest common divisor of n and k . One obtains `gcd(n,k) = 1`. Bob now publishes the numbers m and k .

2. Sender Alice transcribes the message

PRIME NUMBERS ARE USEFUL!

using the ASCII encoding as

$$a = 80827377693278857766698283326582693285836970857633.$$

The SAGE command `map(ord, "PRIME")`, for example, encodes the word "PRIME" in ASCII as `[80,82,73,77,69]`, giving 8082737769. Then Alice computes the integer b such that $b \equiv a^k \pmod{m}$ and $0 < b < m$ and transmits the encrypted message

$$\begin{aligned}
b &= 36057960785874251250398847578656552489665279269698 \\
&\quad 48103809148617096444525775586803496118061034800457 \\
&\quad 72014608577306857911068935474951466578892598687245 \\
&\quad 6073152821301324024745350344439303132600913173384.
\end{aligned}$$

The relevant SAGE command is `power_mod(a,k,m)`, which returns the integer b .

3. To decode the message b , recipient Bob determines the secret key x such that $k \cdot x \equiv 1 \pmod{n}$ and $0 < x < n$. This is accomplished with the SAGE command `k.inverse_mod(n)` (see also Remark B.8). Bob obtains thereby the solution

$$\begin{aligned}
 x = & 10440113532487050550245048433838898684754607388682 \\
 & 99607558952710255800769876309205175474361385321782 \\
 & 50756334845499692617255424236824270979867936841467 \\
 & 99676413130267592026954003935210926047201603331107.
 \end{aligned}$$

Bob now decodes the message b using the SAGE command

$$\text{power_mod}(b, x, m)$$

to determine the unique integer c such that $c \equiv b^x \pmod{m}$ and $0 < c < m$. Bob thereby obtains the number

$$c = 80827377693278857766698283326582693285836970857633,$$

that is, the message “PRIME NUMBERS ARE USEFUL!”

Remark B.19. In applications with limited storage space (chip cards, for example), there is an increased use of asymmetric encryption algorithms that use elliptic curves. Instead of the operations $+$ and \cdot on the integers \mathbb{Z} , a special operation of addition of points on a given elliptic curve is defined. The operation a^k corresponds to k -fold addition of a point to itself. We shall learn about addition on elliptic curves in Appendix C. Cryptography that uses elliptic curves is called, not surprisingly, *elliptic curve cryptography* (ECC). For an elementary introduction to this topic, we refer the reader to [8].

References

- [1] M. W. Baldoni, C. Ciliberto, G. M. Placentini Cattaneo: *Elementary number theory, cryptography and codes*. Translated from the 2006 Italian original by D. A. Gewurz. Springer, Berlin, 2009.
- [2] F. L. Bauer: *Decrypted secrets: methods and maxims of cryptology*. Springer, Berlin Heidelberg New York, 4th edition, 2006.
- [3] J. Buchmann: *Introduction to cryptography*. Springer, Berlin Heidelberg New York, 2nd edition, 2004.
- [4] W. Diffie, M. E. Hellman: *New directions in cryptography*. IEEE Trans. Information Theory **IT-22** (1976), 644–654.
- [5] D. Kahn: *The codebreakers. The comprehensive history of secret communication from ancient times to the internet*. Simon & Schuster, 2nd edition, 1997.
- [6] R. L. Rivest, A. Shamir, L. Adleman: *A method for obtaining digital signatures and public-key cryptosystems*. Comm. ACM **21** (1978), 120–126.
- [7] S. Singh: *The code book: the science of secrecy from ancient Egypt to quantum cryptography*. Random House, 2011.
- [8] L. Washington: *Elliptic curves: number theory and cryptography*. CRC Press, 2nd edition, 2008.

<http://www.springer.com/978-3-319-69427-6>

From Natural Numbers to Quaternions

Kramer, J.; von Pippich, A.-M.

2017, XVIII, 277 p. 10 illus., 6 illus. in color., Softcover

ISBN: 978-3-319-69427-6