Springer Undergraduate Mathematics Series

S
U
M
S

Jürg Kramer
Anna-Maria von Pippich

From Natural Numbers to Quaternions

Springer

Springer Undergraduate Mathematics Series

Advisory Board

M. A. J. Chaplain, *University of St. Andrews*
A. MacIntyre, *Queen Mary University of London*
S. Scott, *King's College London*
N. Snashall, *University of Leicester*
E. Süli, *University of Oxford*
M. R. Tehranchi, *University of Cambridge*
J. F. Toland, *University of Cambridge*

More information about this series at http://www.springer.com/series/3423

Jürg Kramer · Anna-Maria von Pippich

From Natural Numbers to Quaternions

Springer

Jürg Kramer
Department of Mathematics
Humboldt-Universität zu Berlin
Germany

Anna-Maria von Pippich
Department of Mathematics
Technische Universität Darmstadt
Germany

Translation from the German language edition: *Von den natürlichen Zahlen zu den Quaternionen* by Jürg Kramer and Anna-Maria von Pippich, © Springer Spektrum 2013. All Rights Reserved.

ISSN 1615-2085 ISSN 2197-4144 (electronic)
Springer Undergraduate Mathematics Series
ISBN 978-3-319-69427-6 ISBN 978-3-319-69429-0 (eBook)
https://doi.org/10.1007/978-3-319-69429-0

Library of Congress Control Number: 2017958024

Mathematics Subject Classification (2010): 08–01, 11–01, 12–01, 20–01

© Springer International Publishing AG 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface to the English Edition

This book on the construction of number systems first appeared in 2013 in a German edition with the same title. It can be seen from the following preface to that edition that the goal of this book is to present a basic and comprehensive construction of number systems, beginning with the natural numbers and ending with Hamilton's quaternions, while providing relevant algebraic knowledge along the way. As a supplement to the German edition, an appendix has been added to each chapter in this English edition, which in contrast to the rigorous style of the rest of the book, presents in the more casual form of a survey some related aspects of the material of the chapter, including some recent developments.

We would like to offer our most heartfelt thanks to the translator, David Kramer, for his competent work, which has contributed significantly to this English version and in many places led to a more felicitous presentation of the material.

We hope that this book will help students and teachers of mathematics as well as all those with an interest in the subject to fill in any gaps in their mathematical education related to the construction of number systems and that the appendices will inspire some readers to pursue further mathematical studies.

Berlin, September 2017

Jürg Kramer
Anna-Maria von Pippich

Preface to the German Edition

The main topic of this book is an elementary introduction to the construction of the number systems encountered by mathematics students in their first semesters of study. Beginning with the natural numbers, we successively construct, along with the requisite algebraic machinery, all the number fields containing the natural numbers, including the real numbers, complex numbers, and Hamiltonian quaternions. Our experience has shown us that time is frequently lacking in introductory mathematics courses for a well-founded construction of number systems; this book represents a contribution toward filling that gap.

The construction of number systems also represents an important component in the professional education of mathematics teachers. For this reason, this book offers a self-contained and compact construction of the number systems that are of relevance to different grade levels from a mathematical perspective with a view toward aspects of pedagogical content knowledge.

This book arose from a course in elementary abstract algebra and number theory given a number of times at the Humboldt University of Berlin. Parts of the first-named author's book *Zahlen für Einsteiger: Elemente der Algebra und Zahlentheorie* (Vieweg Verlag, Wiesbaden, 2008) have been revised and expanded for inclusion in this newly conceived book on the construction of number systems. Numerous exercises with extensive solutions facilitate the reader's engagement with the subject.

The completion of this book would not have been possible without the contributions of many individuals. Here we wish to thank first of all Christa Dobers and Matthias Fischmann for typing the first parts of the manuscript. In addition, we wish to thank all the students whose written course notes contributed to the text. We also wish to thank our colleagues, in particular Andreas Filler and Wolfgang Schulz, for their numerous suggestions for improving early versions of the manuscript. A special word of thanks goes to Olaf Teschke for his work on creating the exercises, and we also thank Barbara Jung and André Henning for their work on writing up solutions to the exercises. Finally, we offer hearty thanks to Christoph Eyrich for his expert support in designing the layout of the book and to Ulrike Schmickler-Hirzebruch for her encouragement and support on behalf of the publisher, Springer Spektrum.

Berlin, February 2013

Jürg Kramer
Anna-Maria von Pippich

Table of Contents

Introduction

The Development of the Integers and Algebra

One of mankind's earliest intellectual occupations was counting. The development of the concepts of numbers and the representation of numbers has therefore assumed a place of importance in the history of every civilization. The enormous effectiveness of our decimal system of numerical representation is the culmination of centuries—indeed millennia—of earlier efforts that together represent a powerful cultural attainment.

The idea of counting objects, that is, of bringing a set of equivalent objects into a one-to-one correspondence with a fixed set of numbers, represents a significant intellectual process of abstraction.

In more advanced cultures, systems of symbolic notation for these numbers—some more effective than others—were developed. We mention particularly the cuneiform writing of the Babylonians, Egyptian hieroglyphics, Roman numerals, and the system of numerals developed in India. It was only in the thirteenth and fourteenth centuries that the Indian positional decimal system finally made its way via the Islamic world to Western Europe, which to this day uses "Arabic" numerals.

The development of number systems goes relatively closely hand in hand with the development of methods of calculation. In this regard, the Babylonian and Indian number systems, for example, were far superior to those of the Egyptians and Romans. Nevertheless, until late in the fifteenth century, in both the ancient civilizations and Western Europe, numerical calculation was the province of a small group of specialists known as arithmeticians. It was not until the publication in the fifteenth century of Adam Ries's books on calculation, which were based on the book *Liber Abaci* of Leonardo of Pisa, known as Fibonacci, that the usual methods of calculation that we use today became accessible to the "common people." The diffusion of calculational techniques is linked to a systemization of arithmetic in the academic world, which then led to the development of algebra. At first, algebra was viewed primarily as a practical tool, but it gradually took on a life of its own and eventually developed into the independent discipline that we know today. Algebra will therefore play a significant role in every rigorous scientifically based construction of number systems.

A First Look at Number Systems

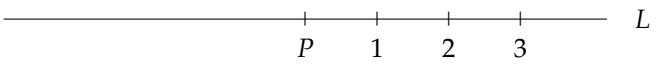
We all recall from our schooldays how we first learned about the numbers $1, 2, 3, \ldots$, then the square roots of such numbers, for example $\sqrt{2}$, and somewhat later became acquainted with the number π, associated with the circumference of a circle, and perhaps Euler's constant e. On our first encounter with these numbers, we had no idea that a powerful intellectual construct had to be developed before a number system could be created that could contain all these numbers and make possible a "sensible" way of calculating with them, namely the system of real numbers. The creation of this number system represents an outstanding achievement of the human intellect, and a fundamental objective of this book is to acquaint students with the construction of the real numbers so that they may become familiar with the fine structure of these objects.

It is astounding that the set of real numbers, which we denote by $\mathbb{R}$, can be developed essentially from the single number 1 (one). Let us sketch briefly how this is done, for a thoroughgoing working out of this process is the main purpose of this book. We begin by identifying the number 1 with an object, and we then bring along another object of the same kind, so that we now have two objects and thereby have acquired the number 2. We may formalize this process by writing $2 = 1 + 1$. Continuing in this manner, we obtain in sequence the numbers

$$\begin{aligned} 3 &= 2 + 1 = 1 + 1 + 1, \\ 4 &= 3 + 1 = 1 + 1 + 1 + 1, \\ &\ldots \end{aligned}$$

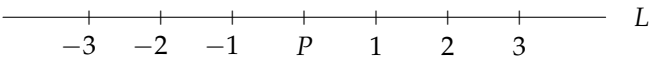
that is, the set of natural numbers $\mathbb{N}$ except for the number 0 (zero), which we shall obtain momentarily, and append to the set of natural numbers. One might say that the number 1 generates additively every natural number. That is, the number 1 is, from an additive point of view, the atom from which every natural number is built.

We may picture the natural numbers $1, 2, 3, \ldots$ as sitting equally spaced like pearls on a necklace beginning at the left with 1 and continuing sequentially off to the right. We might also represent these numbers geometrically, to which end we choose a unit length and mark it off on a horizontal line L by beginning at a point P and moving to the right. We denote by the symbol 1 the point on the line thereby constructed. Continuing, we obtain a second point, which we denote by 2, and so on:



For no reason other than symmetry we might wish to carry out a similar process by moving to the left. Of course, the new points that we thereby

obtain must be given new names. We denote the mirror image of 1 in the point P by -1, and so forth, obtaining



We denote the reflection point P by 0. What we have obtained here in a very graphic way is the process of extending the system of natural numbers $\mathbb{N}$ to the system of integers $\mathbb{Z}$. This can be interpreted algebraically by saying that the equation

$$x + n = m \quad (m, n \in \mathbb{N})$$

always has a solution.

Up to now, we have taken exclusively the additive point of view. We could, however, combine natural numbers and integers in a second way, namely multiplicatively. Just as the addition of numbers can be interpreted as laying off unit distances on a line, multiplication can be viewed as measuring area with respect to the unit square (a square all of whose sides have length equal to the unit distance 1).

If for $n \in \mathbb{N}$, we define

$$n \cdot 0 := 0$$

and then inductively

$$n \cdot (m + 1) := (n \cdot m) + n,$$

we obtain the formal equivalent of this. Just as we recognized 1 as the atom for the additive construction of the natural numbers and integers, we may now ask the corresponding question in the multiplicative case. The answer turns out to be much more complicated: we are led to the (infinite set of) prime numbers. That every integer can be uniquely (up to order and units) represented as the product of prime numbers is not a priori clear. This fact is the content of the fundamental theorem of arithmetic.

From an algebraic point of view, one can, in analogy to the additive case, ask in the multiplicative case about the unrestricted solvability of the equation

$$n \cdot x = m \quad (m, n \in \mathbb{Z}).$$

To be sure, there is no solution when $n = 0$ and $m \neq 0$. But what about the case $n \neq 0$? In the general case, there is no solution $x \in \mathbb{Z}$ unless n is a divisor of m.

To overcome this restriction, one is led to the set $\mathbb{Q}$ of rational numbers. Such numbers are familiar to us as "fractions" $r = \frac{m}{n}$ $(m, n \in \mathbb{Z};\ n \neq 0)$. One should note, however, that the representation of r in the form $\frac{m}{n}$ is not unique: we can make the numerator and denominator larger or smaller using the relation

$$r = \frac{m}{n} = \frac{m'}{n'} \Longleftrightarrow m \cdot n' = n \cdot m'.$$

It is therefore essential in understanding the set $\mathbb{Q}$ that we imagine a rational number as a class of pairs of integers. At this point, there are several ways in which we might motivate a further enlargement of our number system. For example, we could follow the ancient Greeks and use the fundamental theorem of arithmetic to demonstrate that the length of the diagonal of the unit square, that is, the "number" $\sqrt{2}$, is not rational, which would then require an enlargement of the number system $\mathbb{Q}$. Another possibility is the following: using the geometric representation of the integers as points on a line to create an extension to the rational numbers using properties of similar triangles, we obtain these new rational numbers as additional points that are "densely packed" on the line. The question now arises whether these newly obtained points constitute the entire number line, that is, whether the rational numbers fill up the number line without leaving any gaps. As is well known, there are such gaps, and we are then motivated to attempt to fill them in. Once again, one is led to an extension of the set of rational numbers $\mathbb{Q}$ and thereby to the construction of the real numbers $\mathbb{R}$. This nontrivial process of completion of the rational numbers has wide-ranging consequences, since it lays the foundation for calculus and thereby makes it possible, for example, to handle differential equations, which describe many processes in the real world.

Detailed Outline of This Book

There are various aspects of the natural numbers that can provide us an orientation. One relies primarily on the *cardinal aspect* (counting aspect) and the *ordinal aspect* (ordering aspect) of the natural numbers. The cardinal aspect rests on the equivalence classes of sets of equal size, while the ordinal aspect is based on the assumption that the set of natural numbers has a beginning point, that every natural number has exactly one successor number, and that distinct natural numbers have distinct successors. In the framework of our axiomatic approach, we turn to the ordinal aspect and establish the natural numbers at the beginning of the first chapter with the help of Peano's axioms. Using the fifth Peano axiom, namely the axiom of mathematical induction, we define addition and multiplication of the natural numbers and introduce the usual arithmetic operations. In the second part of the first chapter, we develop the concept of divisibility of natural numbers; the main result of this part is the proof of the fundamental theorem of arithmetic. The first chapter ends with a section on division with remainder, in which the decimal representation of numbers plays an important role.

The operations of addition and multiplication of natural numbers developed in the first chapter are abstracted in the second chapter, leading to the

definitions of semigroups and monoids. These notions begin our construction of number systems through the necessary algebraic concepts that we develop in the second and third chapters. In the second chapter, we concentrate above all on an elementary presentation of group theory, introducing groups, subgroups, normal subgroups, group homomorphisms, cosets, and quotient groups. These theoretical considerations lead to the fact that regular abelian semigroups can be extended essentially uniquely to groups. This yields in particular the mathematically based extension of the additive semigroup $(\mathbb{N}, +)$ of natural numbers to the additive group $(\mathbb{Z}, +)$ of integers.

The extension of the multiplication of natural numbers to the newly constructed domain of the integers leads to the algebraic concept of a ring. The study of the fundamental aspects of ring theory is the subject of the third chapter. In this connection, we study rings, subrings, ideals, ring homomorphisms, and quotient rings. We discover the special classes of rings known as integral domains and fields, which again play an important role in the construction of number systems. In fields, for example, division of two elements can be carried out in every case, provided the denominator is not zero. We shall see that every integral domain can be enlarged to a field. Since the ring $(\mathbb{Z}, +, \cdot)$ will turn out to be an integral domain, we will be able to enlarge it to the field $(\mathbb{Q}, +, \cdot)$ of rational numbers. The third chapter closes with a discussion of special rings motivated by an algebraic systemization of the concept of divisibility.

To begin the fourth chapter, we apply the decimal representation of integers to the set of rational numbers that we constructed in the third chapter. We thereby obtain the decimal fraction development of rational numbers. It turns out that such a representation of a rational number either terminates or is periodic. This raises the question whether there exists an extension of the rational numbers that contains all "numbers" that can be represented by an arbitrary decimal expansion. As we shall see, this is the set of real numbers, but we have far to go before we can carry out that construction: with the help of the quotient ring of rational Cauchy sequences modulo the ideal of rational null sequences, we first construct a field that contains $\mathbb{Q}$. We determine that this field is complete, that is, that every Cauchy sequence with elements in this field has a limit in this field. From this, we achieve the insight that this abstractly constructed field can be identified with the set of numbers represented by infinite decimals, which leads to the field $\mathbb{R}$ of real numbers. In the last part of the chapter, we consider alternative characterizations of the completeness of $\mathbb{R}$, such as the existence of the supremum of every subset of $\mathbb{R}$ that is bounded from above. Another important point at the end of this chapter is the identification of $\mathbb{R}$ with the number line, which becomes possible only after the axioms of classical Euclidean geometry are extended by a further axiom that postulates that the number line has, so to speak, no holes.

The fifth chapter begins with the question of a further extension of the set of real numbers: given that the integers and rational numbers were created

with the goal of being able to solve every linear equation of the form

$$a \cdot x + b = c \quad (a, b, c \in \mathbb{N}; a \neq 0),$$

the question naturally arises concerning the solvability of equations of higher degree, for example those of degree 2. With quadratic extensions it becomes clear that the solution of quadratic equations implies the existence of square roots. It turns out that real square roots exist for every positive real number. In contrast, no negative real number has a real square root. By postulating that the number -1 has the imaginary unit i as a square root, we are led to the field $\mathbb{C}$ of complex numbers. Having constructed $\mathbb{C}$, we soon come to the conclusion that extraction of square roots can be carried out without restriction in the complex numbers. That in fact, every polynomial equation with complex coefficients has complex roots is the content of the fundamental theorem of algebra, for which we give an elementary proof. In the second part of the chapter we investigate the fine structure of the real (and complex) numbers. This leads us to the distinction between algebraic and transcendental numbers. Although transcendental numbers seem to be a priori more difficult to deal with, their characterization shows that they are particularly well approximated by rational numbers. The chapter closes with a proof of the transcendence of Euler's number $e = 2.71828\ldots$.

Our goal in the final sixth chapter is to search for fields that extend the field of complex numbers $\mathbb{C}$ to an even more encompassing field. Since we can view $\mathbb{C}$ as a two-dimensional real vector space, it makes sense to begin by looking for a field that arises from a three-dimensional real vector space. It turns out, however, that no such field exists. If we then look for a field that can be obtained from a four-dimensional real vector space, we discover that such a field exists, provided that we abandon the requirement of commutativity of multiplication. In this way, we are led to the construction of the skew field of Hamiltonian quaternions, which brings to a close our investigation of number systems.

Chapter Appendices for the Interested Reader

As mentioned in the preface, each of the six chapters in the English edition on the construction of number systems has been supplemented with an appendix. These appendices have been designed for the reader who wishes to learn about some of the further developments to which the number system presented in the corresponding chapter has given rise, both historically and with respect to very recent results. In contrast to the systematic development of the mathematical machinery necessary for constructing number systems, we have adopted in the appendices a less rigorous style. This allows the appendices to remain largely independent of the remainder of the book, and in particular, they should provide the student some first insights

into questions that are topics of current research. The choice of topics largely represents the authors' personal mathematical taste.

The appendix to the first chapter deals with interesting developments on the subject of prime numbers, including work on conjectures that remain unresolved to this day. The appendix to the second chapter provides an introduction to working with congruences, which are particularly useful in cryptographic applications. Building on that knowledge, we introduce the RSA encryption procedure and discuss some of its strengths and weaknesses. In the appendix to the third chapter we investigate the search for rational solutions of polynomial equations in several variables (with integer coefficients), the most famous example being the Fermat equation $X^d + Y^d = Z^d$, which for exponent $d > 2$ has only the trivial solution. In the fourth chapter, after we have obtained the real numbers by completing the rational numbers with respect to the (archimedean) absolute value, we introduce in the appendix the so-called p-adic completion, which leads to the p-adic numbers $\mathbb{Q}_p$, which in turn are helpful in finding rational solutions to polynomial equations in the context of the local–global principle. Following the construction of the complex numbers in the fifth chapter, we turn naturally to the question of the representation in terms of radicals of the zeros of a polynomial in a single variable (with complex coefficients), which turns out to be impossible in general once the degree of the polynomial exceeds four. This leads directly to Galois theory and the current topic of so-called Galois representations. In the appendix to the last chapter, we conclude our book by asking whether there can exist a number system that extends Hamilton's quaternions. It turns out that if we are willing to give up associativity of multiplication, there is precisely one additional extension, Cayley's octonions, which rounds out the subject of this book in a most satisfying way.

Prerequisites

A first prerequisite for the study of this book is an acquaintance with naive set theory. We assume that the interested reader is familiar with the notions of set, membership, and containment, as well as the operations of set intersection, union, and difference. Furthermore, we assume familiarity with mappings between sets and the notions of injectivity, surjectivity, and bijectivity of mappings. It is only in the fifth and sixth chapters that we invoke in certain places the theory of finite-dimensional vector spaces, and we also make use of elements of the calculus of functions of a single real variable.

Final Remarks

Many interesting topics and mathematical pearls from the theories of elementary number theory and abstract algebra are not mentioned in this book.

We have focused primarily on the construction of number systems and their requisite algebraic apparatus. It is our hope that through the lens of algebra, the reader will obtain new insights into the structure of the number systems studied earlier in school, and in addition, will learn with the help of familiar numbers to value the abstract and fruitful methods of algebra in the spirit of the fifth-century Greek philosopher Proclus, who wrote, "Wherever there is number, there is beauty."

I The Natural Numbers

1. The Peano Axioms

We begin our study of elementary number theory with a discussion of the set of natural numbers. According to Leopold Kronecker, the set of natural numbers $\{0,1,2,\dots\}$ together with the familiar operations of addition and multiplication may be considered as having been created by God. We shall not go any further into metaphysics regarding the natural numbers. Instead, we are going to take an axiomatic approach to constructing sets of numbers, and we shall begin by defining the natural numbers with the help of the axioms proposed by Giuseppe Peano.

Definition 1.1 (Peano axioms). The set $\mathbb{N}$ of *natural numbers* is characterized by the following axioms:

(i) The set $\mathbb{N}$ is nonempty. It contains a distinguished element $0 \in \mathbb{N}$.

(ii) For every $n \in \mathbb{N}$, there exists a uniquely defined element $n^* \in \mathbb{N}$ with $n^* \neq n$. The element n^* is called the *(immediate) successor* of n, and n is called the *(immediate) predecessor* of n^*.

(iii) There is no element $n \in \mathbb{N}$ whose successor n^* satisfies the relation $n^* = 0$. That is, the element 0 has no predecessor and is therefore considered the *first element*.

(iv) If two natural numbers n_1, n_2 satisfy the equality $n_1^* = n_2^*$, then it follows that $n_1 = n_2$. That is, the successor mapping is an injection from $\mathbb{N}$ to $\mathbb{N}$.

(v) *Principle of mathematical induction:* If T is a subset of $\mathbb{N}$ with the property that $0 \in T$ (basis of the induction), and if it follows from the assumption $t \in T$ (induction hypothesis) that $t^* \in T$ as well (induction step), then we must have $T = \mathbb{N}$.

Remark 1.2. Note that *the* successor of a natural number n is the immediate successor n^*. By *a* successor of a natural number n we mean any element of the set $\{n^*, n^{**}, n^{***}, \dots\}$; we will also call these numbers respectively the first successor, second successor, third successor, etc., of n. Similar nomenclature holds for predecessors.

Remark 1.3. Using Definition 1.1 repeatedly, we may introduce the following familiar notation:

$$1 := 0^*, \quad 2 := 1^* = 0^{**}, \quad 3 := 2^* = 1^{**} = 0^{***}, \quad \dots,$$

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0_1

where the multiple asterisks denote multiple applications of taking the successor. The set $\mathbb{N}$ of natural numbers can now be written in the familiar form

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

Axiom (v) of Definition 1.1 forms the basis of constructing what are called *proofs by induction:* if one wishes to prove that every natural number possesses a certain property, then one may do so by first proving that the number 0 possesses the given property (basis of the induction), then showing that on the assumption that the natural number n ($n \in \mathbb{N}$ arbitrary but fixed) has the property (induction hypothesis), it follows that the successor n^* has the property. By Axiom (v), it follows that the property holds for all $n \in \mathbb{N}$.

We would like to note here that the principle of mathematical induction can be formulated in the following modified form: if T is a subset of $\mathbb{N}$ containing n_0 (basis of induction) and if $t \in T$ (induction hypothesis) implies $t^* \in T$ (induction step), then we must have $T \supseteq \{n_0, n_0^*, \dots\}$. With induction proofs of this type, it is possible to prove properties that do not hold necessarily for all the natural numbers, but only for n_0 and its successors.

Remark 1.4. One could justifiably ask whether the set of natural numbers as defined actually exists, whether there exists a *model* of the Peano axioms. This question can be answered in the affirmative with the help of set theory. Likewise, one can show that the natural numbers are uniquely determined, that is, that all models of the Peano axioms are equivalent (more precisely, isomorphic) to each other. We refer the reader to the vast literature on set theory.

We now define the operations of addition and multiplication of natural numbers.

Definition 1.5. *Addition* and *multiplication* of natural numbers m and n are defined inductively as follows:

$$\text{Addition:} \quad n + 0 := n, \quad n + m^* := (n + m)^*, \tag{1}$$

$$\text{Multiplication:} \quad n \cdot 0 := 0, \quad n \cdot m^* := (n \cdot m) + n. \tag{2}$$

Remark 1.6. Definition 1.5 indeed constitutes a valid definition of addition and multiplication on the set of natural numbers. If one wishes, for example, to add the natural number n to the natural number m, the *sum* $n + m$ is determined by (1) as follows: we write m in the form $m = 0^{*\cdots*}$ (with m asterisks; that is, m is the mth successor of 0); from $n + 0 = n$ by definition, we may deduce

$$\begin{aligned} n+1 &= n+0^* = (n+0)^* = n^*, \\ n+2 &= n+0^{**} = (n+0^*)^* = (n+1)^* = n^{**}, \\ &\cdots \\ n+m &= n+0^{*\cdots *} = n^{*\cdots *} \ (m \text{ times}); \end{aligned}$$

that is, the sum $n+m$ is the mth successor of n.

Similarly, the *product* $n \cdot m$ of $n, m \in \mathbb{N}$ is defined by (2). We note here that we will frequently omit the dot symbolizing multiplication and write mn instead of $m \cdot n$.

We shall now use Peano's axioms to prove that the usual laws of addition and multiplication hold under our definition of those two operations.

Lemma 1.7. *The following laws hold for arbitrary natural numbers n, m, p:*

- *Associative law:*

$$\begin{aligned} n+(m+p) &= (n+m)+p, \\ n \cdot (m \cdot p) &= (n \cdot m) \cdot p. \end{aligned}$$

- *Commutative law:*

$$\begin{aligned} n+m &= m+n, \\ n \cdot m &= m \cdot n. \end{aligned}$$

- *Distributive law:*

$$\begin{aligned} (n+m) \cdot p &= (n \cdot p)+(m \cdot p), \\ p \cdot (n+m) &= (p \cdot n)+(p \cdot m). \end{aligned}$$

Proof. We shall present a proof for the commutative law of addition. The other proofs are similar. We shall use a double induction argument, that is, an induction on m, and embedded within that induction, an induction on n.

(i) We take $m = 0$ as the basis of the induction: We must show that

$$n+0 = 0+n$$

for all $n \in \mathbb{N}$. Since by (1), we have $n+0 = n$, we must show that $0+n = n$; we do this by induction on n. For $n = 0$, the assertion is true. Given the induction hypothesis that $0+n = n$ for an arbitrary $n \in \mathbb{N}$, we must show that $0+n^* = n^*$. Using (1) and the induction hypothesis, we see easily that

$$0+n^* = (0+n)^* = n^*.$$

This completes the induction on n and also the basis of the induction for $m = 0$.

(ii) We now make the induction hypothesis that for $m \in \mathbb{N}$, the equality

$$n + m = m + n$$

holds for all $n \in \mathbb{N}$. On this assumption, we now assert that we also must have

$$n + m^* = m^* + n$$

for all $n \in \mathbb{N}$. Before we prove this, we show first that $m^* + n = (m + n)^*$ for all $n \in \mathbb{N}$, again using induction on n. For $n = 0$, the assertion follows at once from (1). With the induction hypothesis $m^* + n = (m + n)^*$, it suffices to show that we also have $m^* + n^* = (m + n^*)^*$. This can be seen at once from using (1) twice and the induction hypothesis, namely

$$m^* + n^* = (m^* + n)^* = \big((m + n)^*\big)^* = (m + n^*)^*.$$

This allows us to complete the induction on m. Namely, using (1), the induction hypothesis, and the equality that we just proved, we have

$$n + m^* = (n + m)^* = (m + n)^* = m^* + n.$$

This completes the proof of the fact that addition of natural numbers is commutative. □

Exercise 1.8. Prove the remaining laws of addition and multiplication from Lemma 1.7.

Remark 1.9. In connection with the distributive law, we note that the operation of multiplication takes precedence over that of addition. Thus, recalling that we may suppress the dot symbolizing multiplication, we can write the distributive law in the following form:

$$(n + m)p = np + mp,$$
$$p(n + m) = pn + pm.$$

Exercise 1.10. Prove the following assertion: The product of two natural numbers m and n is equal to 0 if and only if at least one of the two numbers is equal to 0.

Remark 1.11. In defining addition and multiplication of natural numbers we have assumed the validity of the principle that it is possible to define functions on the natural numbers recursively. This means that in defining a function f on $\mathbb{N}$, it suffices to define the value $f(0)$ and then to define $f(m^*)$ in terms of m and $f(m)$. For a proof of this principle we refer the reader, for example, to Section 2.10 of the book "Set theory: with an introduction to real points sets", by A. Dasgupta.

To simplify notation, we now introduce exponential notation.

Definition 1.12. Let a and m be two natural numbers. We define the *mth power a^m of a* inductively on m as follows:

$$a^0 := 1,$$
$$a^{m^*} := a^m \cdot a.$$

Lemma 1.13. *Let a, m, n be arbitrary natural numbers. Then we have the following rules:*

$$a^m \cdot a^n = a^{m+n},$$
$$(a^m)^n = a^{m \cdot n}.$$

Proof. We leave the proof as an exercise for the reader. □

Exercise 1.14. Prove the power law of Lemma 1.13.

Definition 1.15. Le $m, n \in \mathbb{N}$ be given. We say that m is *less than or equal to n*, and write

$$m \leq n,$$

if either m is a predecessor of n or $m = n$. If $m = n$ does not hold, then we say that m is *(strictly) less than n* and write

$$m < n.$$

We define analogously the notion of m being *greater than or equal to n*, and write

$$m \geq n,$$

if either m is a successor of n or $m = n$. If the equality $m = n$ does not hold, then we say that m is *(strictly) greater than n*, and we write

$$m > n.$$

Remark 1.16. With the relation $<$, the set of natural numbers $\mathbb{N}$ becomes an *ordered set*; that is, the following three conditions are satisfied:
(i) For all elements $m, n \in \mathbb{N}$, 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 conditions hold for the relation $>$.

Exercise 1.17. Prove properties (i), (ii), and (iii) of Remark 1.16.

Remark 1.18. Using the relation $<$, we can present the following variant of proof by induction, also called *strong induction:* Suppose we wish to prove that every natural number $n \geq n_0$ satisfies a certain property. Then we first

show that the natural number n_0 possesses this property (basis of the induction), then select an arbitrary natural number $n > n_0$ and assume that the property in question holds for every natural number n' such that $n_0 \leq n' < n$ (induction hypothesis), and show finally that on the induction hypothesis, the natural number n also possesses this property (induction step).

Remark 1.19. For the relation $<$, we have the following rules relating to addition and multiplication:
(i) For all $p \in \mathbb{N}$, if $m < n$, then also $m + p < n + p$.
(ii) For all $p \in \mathbb{N}$, $p \neq 0$, if $m < n$, then also $m \cdot p < n \cdot p$.
Analogous rules hold for the relation $>$.

Exercise 1.20. Prove properties (i) and (ii) of Remark 1.19.

Lemma 1.21 (Well-ordering principle). *If $M \subseteq \mathbb{N}$ is a nonempty subset of the natural numbers, then M contains a smallest element m_0. That is, for all $m \in M$, we have $m \geq m_0$.*

Proof. Let m be an arbitrary fixed element of M. We provisionally set $m_0 := m$. By the ordering of M (see Remark 1.16), m_0 can be compared with any element of M, and it can thus be decided whether m_0 has a predecessor in M. If not, then m_0 is the desired least element, and we are done. If, however, m_0 has a predecessor $m' \in M$, that is, $m'^{*\cdots*} = m_0$, then we reset $m_0 := m'$. We again ask whether m_0 has a predecessor in M. If not, we are done. Otherwise, we proceed as above and find a predecessor of m_0. The possibility of choosing another predecessor must end after at most m steps, since we eventually would reach the first element, 0, of $\mathbb{N}$, which has no predecessor. If $0 \notin M$, the proof will end in fewer steps. □

Remark 1.22. The well-ordering principle ensures the existence of a smallest element in a nonempty set of natural numbers. This does not mean that it is always easy to determine such a minimal element.

For example, it has been proven that there is a (very large) natural number m_1 such that all natural numbers $m \geq m_1$ can be written as a sum of at most seven third powers (cubes). By the well-ordering principle, there must be a least natural number m_0 with this property (the value of m_0 is conjectured to be 455). However, to this day, the true value of m_0 is unknown.

Exercise 1.23. Can you find any examples from everyday life in which a smallest element of a finite set must exist, yet the actual value of that number is impossible to determine in practice?

Definition 1.24. Suppose we have $m, n \in \mathbb{N}$ with $m \geq n$. Then $(m - n)$, or $m - n$ for short, denotes the natural number that satisfies the equation $n + x = m$. We call $m - n$ the *difference of m and n*.

Exercise 1.25. Prove that the difference $m - n$ of two numbers $m, n \in \mathbb{N}$ with $m \geq n$ is well defined, that is, that there exists precisely one natural number x that satisfies the equation $n + x = m$.

Remark 1.26. One motivation for including the set of natural numbers in a larger set is the desire to have a solution to the equation

$$n + x = m$$

for given natural numbers m, n. Definition 1.24 assumes (and you proved it in the exercise) that a solution exists in the set of natural numbers, namely the number $x = m - n$, when $m \geq n$. Indeed, x is determined by the fact that m is the $x = (m - n)$-fold successor of n. On the other hand, if $m < n$, then there is no natural number that can be substituted for x that will solve the equation. This deficit will lead us to the construction of the *integers*, which we shall be able to accomplish with some algebraic tools to be presented in the next chapter.

2. Divisibility and Prime Numbers

We begin with the definition of divisibility in the natural numbers.

Definition 2.1. A natural number $b \neq 0$ *divides* a natural number a, denoted by $b \mid a$, if there exists a natural number c such that $a = b \cdot c$. We say also that b is a *divisor* of a. We say that $b \in \mathbb{N}$ is a *common divisor* of $a_1, a_2 \in \mathbb{N}$ if there exist $c_1, c_2 \in \mathbb{N}$ such that $a_j = b \cdot c_j$ for $j = 1, 2$.

Example 2.2. Let $a = 12$ and $b = 6$. Then with $c = 2$, the equation $a = b \cdot c$ is satisfied; therefore, $6 \mid 12$. On the other hand, if we take $a = 12$ and $b = 7$, then $7 \nmid 12$.

If we take $a_1 = 12$, $a_2 = 6$, and $b = 3$, then we can see that 3 is a common divisor of 12 and 6.

Remark 2.3. Let a be a nonzero natural number, and b a divisor of a (that is, $a = b \cdot c$ for some $c \in \mathbb{N}$, $c \geq 1$) such that $a \neq b$. Then we must have $b < a$. Indeed, if we had $b > a$, we would be led by Remark 1.19 to the inequality

$$a = b \cdot c \geq b \cdot 1 = b > a,$$

which is impossible.

We may conclude at once from this discussion that if the equation $m \cdot n = 1$ is satisfied by natural numbers, then we must have $m = n = 1$. Namely, we have $m \mid 1$, and from the assumption $m \neq 1$, we would have by the above that $m < 1$, that is, $m = 0$, which is impossible, since that would lead to the equality $0 = 1$.

Lemma 2.4. *We have the following basic facts about divisibility in the natural numbers:*
(i) $a \mid a \quad (a \in \mathbb{N}; a \neq 0)$.
(ii) $a \mid 0 \quad (a \in \mathbb{N}; a \neq 0)$.
(iii) $1 \mid a \quad (a \in \mathbb{N})$.
(iv) $c \mid b,\ b \mid a \Rightarrow c \mid a \quad (a, b, c \in \mathbb{N}; b, c \neq 0)$.
(v) $b \mid a \Rightarrow b \cdot c \mid a \cdot c \quad (a, b, c \in \mathbb{N}; b, c \neq 0)$.
(vi) $b \cdot c \mid a \cdot c \Rightarrow b \mid a \quad (a, b, c \in \mathbb{N}; b, c \neq 0)$.
(vii) $b_1 \mid a_1,\ b_2 \mid a_2 \Rightarrow b_1 \cdot b_2 \mid a_1 \cdot a_2 \quad (a_1, a_2, b_1, b_2 \in \mathbb{N}; b_1, b_2 \neq 0)$.
(viii) $b \mid a_1,\ b \mid a_2 \Rightarrow b \mid (c_1 \cdot a_1 + c_2 \cdot a_2) \quad (a_1, a_2, c_1, c_2, b \in \mathbb{N}; b \neq 0)$.
(ix) $b \mid a \Rightarrow b \mid a \cdot c \quad (a, b, c \in \mathbb{N}; b \neq 0)$.
(x) $b \mid a,\ a \mid b \Rightarrow a = b \quad (a, b \in \mathbb{N}; a, b \neq 0)$.

Proof. Since divisibility properties are of great importance in elementary number theory, we shall present the proofs in detail, even though they are quite straightforward.

(i) By the definition (2) of multiplication of natural numbers, we have for all $a \in \mathbb{N}$ the equality $a = a \cdot 1$. That is, we have $a \mid a$ for $a \neq 0$.

(ii) Likewise, from (2), we have for all $a \in \mathbb{N}$ the relation $0 = a \cdot 0$. That is, we have $a \mid 0$ for $a \neq 0$.

(iii) Using the equality in (i) above and the commutativity of multiplication, we have $a = 1 \cdot a$, from which we obtain $1 \mid a$.

(iv) Since by assumption, we have $c \mid b$ and $b \mid a$, there exist $m, n \in \mathbb{N}$ such that $b = c \cdot m$ and $a = b \cdot n$. We thereby obtain

$$a = b \cdot n = (c \cdot m) \cdot n = c \cdot (m \cdot n),$$

and therefore $c \mid a$.

(v) It follows from $b \mid a$ that there exists $m \in \mathbb{N}$ such that $a = b \cdot m$. On multiplying this equality by $c \in \mathbb{N}$, $c \neq 0$, we obtain the equality $a \cdot c = (b \cdot m) \cdot c$. Taking into account the commutativity and associativity of multiplication, we have $a \cdot c = (b \cdot c) \cdot m$. That is, $b \cdot c \mid a \cdot c$.

(vi) From $b \cdot c \mid a \cdot c$, it follows that there exists $m \in \mathbb{N}$ such that $a \cdot c = (b \cdot c) \cdot m$. As the difference of the left-hand and right-hand sides of this equality, we obtain, using the properties of addition and multiplication for the natural numbers, in particular the distributive property, the equality

$$0 = a \cdot c - (b \cdot c) \cdot m = (a - b \cdot m) \cdot c.$$

But since $c \neq 0$ and the product of $a - b \cdot m$ and c is equal to 0, we must have $a - b \cdot m = 0$, whence $a = b \cdot m$, from which follows $b \mid a$.

(vii) By assumption, there exist $m_1, m_2 \in \mathbb{N}$ such that $a_1 = b_1 \cdot m_1$ and $a_2 = b_2 \cdot m_2$. We thereby obtain, using the properties of addition and multiplication,

$$a_1 \cdot a_2 = (b_1 \cdot m_1) \cdot (b_2 \cdot m_2) = (b_1 \cdot b_2) \cdot (m_1 \cdot m_2),$$

and consequently, $b_1 \cdot b_2 \mid a_1 \cdot a_2$.

(viii) If the number b divides two natural numbers a_1, a_2, then there exist $m_1, m_2 \in \mathbb{N}$ such that $a_1 = b \cdot m_1$ and $a_2 = b \cdot m_2$. Let $c_1, c_2 \in \mathbb{N}$ be arbitrary. For the natural number $c_1 \cdot a_1 + c_2 \cdot a_2$, we obtain by substitution, after a brief calculation,

$$c_1 \cdot a_1 + c_2 \cdot a_2 = c_1 \cdot (b \cdot m_1) + c_2 \cdot (b \cdot m_2) = b \cdot (c_1 \cdot m_1 + c_2 \cdot m_2),$$

from which we conclude that $b \mid (c_1 \cdot a_1 + c_2 \cdot a_2)$.

(ix) Since $b \mid a$, there exists $m \in \mathbb{N}$ such that $a = b \cdot m$. If we multiply this equality by some $c \in \mathbb{N}$, we obtain $a \cdot c = b \cdot (m \cdot c)$, from which $b \mid a \cdot c$ follows at once.

(x) By the divisibility assumptions, both a and b are nonzero. Since $b \mid a$ and $a \mid b$, there exist $n \in \mathbb{N}$ and $m \in \mathbb{N}$ such that $a = b \cdot m$ and $b = a \cdot n$. Substituting the second equality into the first yields

$$a = (a \cdot n) \cdot m \quad \Longleftrightarrow \quad a \cdot (m \cdot n - 1) = 0.$$

Since $a \neq 0$, it follows that $m \cdot n - 1 = 0$; that is, $m \cdot n = 1$. Remark 2.3 tells us at once that $n = m = 1$, from which we conclude that $a = b$.

This completes the proof of the lemma. □

Exercise 2.5. Let $a_1, \ldots, a_k$ be natural numbers such that $a_1 \cdots a_k + 1$ is divisible by 3.
(a) Show that none of the numbers $a_1, \ldots, a_k$ is divisible by 3.
(b) Prove that at least one of the numbers $a_1 + 1, \ldots, a_k + 1$ is divisible by 3.

Remark 2.6. By Lemma 2.4, every $a \in \mathbb{N}$, $a \neq 0$, has the divisors 1 and a. We call these the *trivial divisors of a* . The divisors of $a \in \mathbb{N}$ other than a itself are called *proper divisors of a*.

One can say that from an additive viewpoint, the number 1 is the fundamental building block of the natural numbers, since every natural number can be expressed as a sum of ones. We now consider the multiplicative point of view and ask what might be the fundamental multiplicative building blocks of the natural numbers. This leads us to the notion of prime number, which we now present.

Definition 2.7. A natural number $p > 1$ is called a *prime number* if p has no nontrivial divisors, that is, if p has only the divisors 1 and p. We shall denote the set of prime numbers by

$$\mathbb{P} := \{p \in \mathbb{N} \mid p \text{ is a prime number}\}.$$

Example 2.8. Let us ask whether the number 11 is prime. To this end, we write down all the divisors b of 11. By our considerations above, we must have

$$b \in \{1, \ldots, 11\}.$$

A direct calculation tells us that the numbers $2, \ldots, 10$ cannot be divisors of 11. Therefore, 11 has only the trivial divisors 1 and 11, from which it follows that $11 \in \mathbb{P}$.

The sequence of prime numbers begins

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, \ldots .$$

Lemma 2.9. *Every natural number $a > 1$ has at least one prime divisor $p \in \mathbb{P}$. That is, there exists a prime number p such that $p \mid a$.*

Proof. We consider the following set, whose elements depend on the value of a:

$$\mathcal{T}(a) := \{b \in \mathbb{N} \mid b > 1 \text{ and } b \mid a\}.$$

Since $a \in \mathcal{T}(a)$, the set $\mathcal{T}(a)$ is not empty. By the well-ordering principle (Lemma 1.21), the fact that $\mathcal{T}(a)$ is a nonempty subset of $\mathbb{N}$ implies that it has a least element, which we denote by p. By the way the set was defined, we must have $p > 1$.

We now show that p is a prime number. If such were not the case, that is, if p were not prime, then p would have a proper divisor q greater than 1, that is, we would have $q \mid p$ for some $1 < q < p$. Since $q \mid p$ and $p \mid a$, we obtain from Lemma 2.4 (iv) that $q \mid a$. Since we also have $q > 1$, it follows that $q \in \mathcal{T}(a)$. This contradicts the minimal choice of p. Therefore, p is, as claimed, a prime divisor of a. □

Theorem 2.10 (Theorem of Euclid). *There are infinitely many prime numbers.*

Proof. Suppose such were not the case, that there were only finitely many prime numbers $p_1, \ldots, p_n$. We could then consider the natural number

$$a := p_1 \cdots p_n + 1 .$$

We have certainly $a > 1$, and by Lemma 2.9, a has at least one prime divisor; call it p. By the assumption that there are only finitely many prime numbers, we must have $p \in \{p_1, \ldots, p_n\}$. In particular, $p \mid (p_1 \cdots p_n)$. However, since we also have the divisibility relation $p \mid a$, it follows by the laws of divisibility that $p \mid 1$. That implies that $p = 1$, which, however, is impossible. This refutes the hypothesis that there are only finitely many prime numbers, and so there must be infinitely many primes. □

Remark 2.11. The proof of Euclid's theorem provides us a way of constructing an infinite sequence of prime numbers: We begin with the prime number $p_1 = 2$. Setting $a_2 = p_1 + 1 = 3$, we obtain a second prime, $p_2 = 3$. Setting $a_3 = p_1 \cdot p_2 + 1 = 7$, we obtain the additional prime number $p_3 = 7$. We now

set $a_4 = p_1 \cdot p_2 \cdot p_3 + 1 = 43$ and obtain the prime number $p_4 = 43$. Proceeding in this way, we obtain next $a_5 = p_1 \cdot p_2 \cdot p_3 \cdot p_4 + 1 = 1807$. For the first time in this process, we obtain a number that is not prime, since we have the decomposition $1807 = 13 \cdot 139$. That is, we obtain the two additional prime numbers 13 und 139.

Exercise 2.12. Show that using the procedure of Remark 2.11, one does not obtain every prime number. To accomplish this, define the numbers $a_1 := 2$ and $a_{n+1} := (a_n - 1) \cdot a_n + 1$ ($n \in \mathbb{N}$, $n \geq 1$) and consider the set of prime numbers

$$\mathcal{M}_n := \{p \in \mathbb{P} \mid p \mid a_n\} \qquad (n \in \mathbb{N}, n \geq 1).$$

Show that $\bigcup_{n\in\mathbb{N},n\geq 1} \mathcal{M}_n \neq \mathbb{P}$, by proving that $5 \notin \mathcal{M}_n$ for all $n \in \mathbb{N}$, $n \geq 1$.

Exercise 2.13. Use the idea of the proof of Euclid's theorem and Exercise 2.5 to show that there are infinitely many prime numbers in the subset of the natural numbers

$$2 + 3 \cdot \mathbb{N} := \{2,\ 2+3,\ 2+6,\ \ldots,\ 2+3\cdot n,\ \ldots\}.$$

Example 2.14. We introduce here two special types of prime numbers.

(i) A prime number of the form $p = 2^n - 1$ ($n \in \mathbb{N}$) is called a *Mersenne prime* (after Marin Mersenne). It is known that

$$2^n - 1 \text{ is a prime number} \quad \Longrightarrow \quad n \text{ is a prime number.}$$

(ii) A prime number of the form $p = 2^n + 1$ ($n \in \mathbb{N}$, $n \geq 1$) is called a *Fermat prime* (after Pierre Fermat). It is known that

$$2^n + 1 \text{ is a prime number} \quad \Longrightarrow \quad n = 2^m \text{ for some } m \in \mathbb{N}.$$

Exercise 2.15. Prove the two assertions of Example 2.14.

The converse assertions to (i) and (ii) are in general false. For the converse to (ii), for example, we see that

$$\begin{aligned}
m = 0 &: 2^{2^0} + 1 = 2^1 + 1 \ = 3, && \text{prime,}\\
m = 1 &: 2^{2^1} + 1 = 2^2 + 1 \ = 5, && \text{prime,}\\
m = 2 &: 2^{2^2} + 1 = 2^4 + 1 \ = 17, && \text{prime,}\\
m = 3 &: 2^{2^3} + 1 = 2^8 + 1 \ = 257, && \text{prime,}\\
m = 4 &: 2^{2^4} + 1 = 2^{16} + 1 = 65537, && \text{prime,}
\end{aligned}$$

but the number $2^{2^5} + 1 = 4294967297$ is not prime, since it has the nontrivial divisor 641.

We note here that Carl Friedrich Gauss showed that the regular p-gon ($p \in \mathbb{P}$) can be constructed with straightedge and compass if and only if p is a Fermat prime, that is, a prime of the form $p = 2^{2^m} + 1$ ($m \in \mathbb{N}$).

Example 2.16. A natural number n is said to be *perfect* if the sum of all its divisors is equal to $2n$, that is, if

$$\sum_{d|n} d = 2n.$$

In the first century, the Greek mathematician Nicomachus of Gerasa published a list of the first four perfect numbers: 6, 28, 496, and 8128. The mysterious nature of the perfect numbers has cast its spell over many mathematicians, including Euclid, Mersenne, and Leonhard Euler. All perfect numbers found to date are even numbers. Yet it is unknown whether any odd perfect numbers exist. For even perfect numbers, we can give the following characterization, due to Euler.

Lemma 2.17. *A natural number n is an even perfect number if and only if $n = 2^m \cdot (2^{m+1} - 1)$ for some $m \in \mathbb{N}$ such that $2^{m+1} - 1$ is a prime number.*

Proof. We begin the proof with the following observation: For $n \in \mathbb{N}$, set $S(n) := \sum_{d|n} d$. It is easy to see that for natural numbers a, b whose only common divisor is 1, we have the relationship

$$S(a \cdot b) = S(a) \cdot S(b).$$

Exercise 2.18. Prove this assertion.

With this result, we can now attack the proof. Let $n \in \mathbb{N}$ be an even perfect number. Since n is even, there exist a natural number $m > 0$ and an odd natural number b such that

$$n = 2^m \cdot b.$$

Since n is perfect, our introductory observation yields that $S(n) = S(2^m \cdot b) = S(2^m) \cdot S(b) = 2n$. Since

$$S(2^m) = 2^0 + 2^1 + 2^2 + \cdots + 2^m = \frac{2^{m+1} - 1}{2 - 1} = 2^{m+1} - 1,$$

we obtain the equality

$$(2^{m+1} - 1) \cdot S(b) = 2^{m+1} \cdot b. \tag{3}$$

Therefore, the number $2^{m+1} - 1$ must be a divisor of $2^{m+1} \cdot b$. Using the following exercise, we have in fact that $2^{m+1} - 1$ must divide the number b.

Exercise 2.19. Show that if an odd number $d \in \mathbb{N}$ divides the number $2^{m+1} \cdot b$ $(m, b \in \mathbb{N})$, then d is a divisor of b.

Therefore, there exists $a \in \mathbb{N}$, $a \neq 0$, such that $b = (2^{m+1} - 1) \cdot a$. It remains to show that $a = 1$ and that $2^{m+1} - 1$ is prime.

To this end, we assume that $a > 1$ and show that such an assumption leads to a contradiction. Since $b = (2^{m+1} - 1) \cdot a$, the number b has at least the divisors $\{1, (2^{m+1} - 1), a, b\}$; we therefore have the inequality

$$S(b) \geq 1 + (2^{m+1} - 1) + a + b = 2^{m+1} + a + b = 2^{m+1} \cdot (a + 1).$$

Multiplication by $2^{m+1} - 1$ yields the further inequality

$$\begin{aligned}(2^{m+1} - 1) \cdot S(b) &\geq (2^{m+1} - 1) \cdot 2^{m+1} \cdot (a + 1) > 2^{m+1} \cdot (2^{m+1} - 1) \cdot a \\ &= 2^{m+1} \cdot b,\end{aligned}$$

which contradicts (3). We must therefore have $a = 1$ and $b = 2^{m+1} - 1$. From (3), we conclude that

$$S(b) = 2^{m+1} = b + 1.$$

That is, b has only the divisors 1 and b, and $b = 2^{m+1} - 1$ is therefore a prime number. As asserted, we obtain

$$n = 2^m \cdot (2^{m+1} - 1),$$

with $2^{m+1} - 1$ prime.

We now prove the converse of the statement that we have just proved. Let $n = 2^m \cdot (2^{m+1} - 1)$, where $2^{m+1} - 1$ is prime. From our initial observation, we have that

$$\begin{aligned}S(n) &= S(2^m) \cdot S(2^{m+1} - 1) = (2^{m+1} - 1) \cdot (2^{m+1} - 1 + 1) \\ &= 2 \cdot 2^m \cdot (2^{m+1} - 1) = 2n.\end{aligned}$$

Therefore, n is an even perfect number. □

Exercise 2.20. *(Amicable numbers).* Closely related to the perfect numbers are the *amicable numbers*. Two distinct natural numbers a and b are said to be amicable if $S(a) = a + b = S(b)$; that is, the number a is equal to the sum of the divisors of b that are less than b, and the number b is equal to the sum of the divisors of a that are less than a.

(a) Verify that the numbers 220 and 284 are amicable. This pair was known by the Pythagoreans, as early as 500 B.C..

(b) Prove the following theorem of the Arab mathematician Thabit ibn Qurra: For a fixed natural number n, let us set $x = 3 \cdot 2^n - 1$, $y = 3 \cdot 2^{n-1} - 1$, $z = 9 \cdot 2^{2n-1} - 1$. If x, y, and z are prime, then both $a = 2^n \cdot x \cdot y$ and $b = 2^n \cdot z$ are amicable.

3. The Fundamental Theorem of Arithmetic

We come now to the formulation and proof of the fundamental theorem of arithmetic, which states that the prime numbers are the (multiplicative) building blocks of the natural numbers.

Theorem 3.1 (Fundamental theorem of arithmetic). *Every nonzero natural number a has a representation of the form*

$$a = p_1^{a_1} \cdots p_r^{a_r}$$

as a product of r ($r \in \mathbb{N}$) prime powers of distinct prime numbers $p_1, \ldots, p_r$ with positive natural numbers $a_1, \ldots, a_r$ as exponents. This representation is unique up to the order of the factors.

Proof. We first prove the existence and then the uniqueness of the claimed representation, both by induction on a.

Existence: For $a = 1$, the assertion is true with $r = 0$ (empty product). This establishes the basis of the induction. We now consider $a \in \mathbb{N}$ with $a > 1$, and we take as our induction hypothesis that there exists a prime factorization for every natural number a' with $1 \leq a' < a$. On this assumption, we now prove that a also has a prime factorization. Since we have $a \in \mathbb{N}$, $a > 1$, we know by Lemma 2.9 that a has a prime divisor p. That is, we have

$$a = p \cdot b$$

for some natural number b. Since $p > 1$, it follows that $1 \leq b < a$. By our induction hypothesis, there exists a prime factorization for b of the form

$$b = q_1^{b_1} \cdots q_s^{b_s},$$

where $q_1, \ldots, q_s$ ($s \in \mathbb{N}$) are distinct primes, and $b_1, \ldots, b_s$ are positive natural numbers. Putting everything together, we obtain

$$a = p \cdot b = p^1 \cdot q_1^{b_1} \cdots q_s^{b_s}.$$

If it happens that $p = q_j$ for some $j \in \{1, \ldots, s\}$, we can write the result as

$$a = q_1^{b_1} \cdots q_j^{b_j+1} \cdots q_s^{b_s}.$$

This completes the proof of the existence of a prime factorization for every positive natural number.

Uniqueness: We again employ a proof by induction. As in the existence proof, we begin the induction with $a = 1$ and obtain the uniqueness of the prime factorization of 1 by the fact that the empty product is defined uniquely. We now choose some natural number $a > 1$ and make the induc-

tion hypothesis that the uniqueness of factorization (up to the order of factors) holds for all natural numbers a' with $1 \leq a' < a$. On this assumption, we shall now prove that the prime factorization of a is unique.

In order to achieve a contradiction, we assume that a has two distinct prime factorizations:

$$a = p_1^{a_1} \cdot p_2^{a_2} \cdots p_r^{a_r} = p_1 \cdot b \text{ with } b = p_1^{a_1 - 1} \cdot p_2^{a_2} \cdots p_r^{a_r},$$
$$a = q_1^{b_1} \cdot q_2^{b_2} \cdots q_s^{b_s} = q_1 \cdot c \quad \text{with } c = q_1^{b_1 - 1} \cdot q_2^{b_2} \cdots q_s^{b_s},$$

where r and s are nonzero natural numbers, $p_1, \ldots, p_r$ and $q_1, \ldots, q_s$ are each sets of distinct primes, where we may assume that p_1 is distinct from $q_1, \ldots, q_s$ (why?), and $a_1, \ldots, a_r$ and $b_1, \ldots, b_s$ are all nonzero natural numbers. Without loss of generality, we may also assume that $p_1 < q_1$. Then $a \geq p_1 \cdot c$, and by subtraction, we obtain the natural number

$$a' = a - p_1 \cdot c = \begin{cases} p_1 \cdot (b - c), \\ (q_1 - p_1) \cdot c, \end{cases}$$

for which we have $a' < a$ by construction. The factors $b - c$, $q_1 - p_1$, and c of a' are all natural numbers that are strictly less than a. By the induction hypothesis, the natural numbers a', $b - c$, $q_1 - p_1$, c have each a unique prime factorization. The equality $a' = p_1 \cdot (b - c)$ shows that the prime number p_1 must appear in the prime factorization of a'. The equality $a' = (q_1 - p_1) \cdot c$ shows further that p_1 must appear in the prime factorization of $q_1 - p_1$ or of c. By our assumption, however, p_1 does not appear in the prime factorization of c, so that p_1 must appear in the prime factorization of the difference $q_1 - p_1$. That is, we must have $p_1 \mid (q_1 - p_1)$. Putting this together with $p_1 \mid p_1$ and the equality $q_1 = (q_1 - p_1) + p_1$, we obtain with the help of the divisibility properties that $p_1 \mid q_1$. Since $1 < p_1 < q_1$, we must have that p_1 is a nontrivial divisor of q_1, which is, of course, impossible. We have obtained the desired contradiction. Therefore, our assumption that a has two distinct prime factorizations must also be false. We conclude that the prime factorization of a is unique, which completes the induction step and the proof of uniqueness. □

Exercise 3.2. Find the prime factorization of the following numbers: 720, 9797, 360^{360}, and $2^{32} - 1$.

Invoking the fundamental theorem of arithmetic, we can easily prove the following lemma, which goes back to Euclid.

Lemma 3.3 (Euclid's lemma). *Let a, b be natural numbers and p a prime. Then $p \mid a \cdot b$ implies $p \mid a$ or $p \mid b$.*

Proof. By the assumed divisibility relationship $p \mid a \cdot b$, there exists a natural number $c \neq 0$ such that $a \cdot b = p \cdot c$. Because of the existence and uniqueness of prime factorization, the prime p must appear in the prime factorization of the product $a \cdot b$. Therefore, p must appear in the prime factorization of a or b. This implies at once that $p \mid a$ or $p \mid b$. □

Remark 3.4. By the fundamental theorem of arithmetic, every natural number $a \neq 0$ can be written in the form

$$a = \prod_{p \in \mathbb{P}} p^{a_p},$$

where the product is taken over all prime numbers; we note that only *finitely* many of the exponents a_p are different from 0. We shall formally subsume the case $a = 0$, thus far excluded, in this notation by setting $a_p = \infty$ for all $p \in \mathbb{P}$.

As an application of the fundamental theorem of arithmetic, we derive the following useful divisibility criterion.

Lemma 3.5. *Let a, b be natural numbers with prime factorizations*

$$a = \prod_{p \in \mathbb{P}} p^{a_p}, \quad b = \prod_{p \in \mathbb{P}} p^{b_p}.$$

We then have

$$b \mid a \quad \Longleftrightarrow \quad b_p \leq a_p \quad \text{for all } p \in \mathbb{P}.$$

Remark 3.6. Note that this divisibility criterion is applicable to the case in which $a = 0$ or $b = 0$.

Proof. If b is a divisor of a, then there exists a natural number $c \neq 0$ with $a = b \cdot c$. With the prime factorization

$$c = \prod_{p \in \mathbb{P}} p^{c_p}$$

of c, we obtain

$$\prod_{p \in \mathbb{P}} p^{a_p} = \prod_{p \in \mathbb{P}} p^{b_p} \cdot \prod_{p \in \mathbb{P}} p^{c_p} = \prod_{p \in \mathbb{P}} p^{b_p + c_p}.$$

This proves the equality $a_p = b_p + c_p$, from which follows $b_p \leq a_p$ for all $p \in \mathbb{P}$. The proof of the converse is just as easy. □

Exercise 3.7. Using the criterion of Lemma 3.5, prove that 255 is a divisor of $2^{32} - 1$.

4. Greatest Common Divisor, Least Common Multiple

We begin with the definition of the greatest common divisor.

Definition 4.1. Let a, b be natural numbers, not both equal to 0. A natural number d with the following two properties is called the *greatest common divisor* of 1 and b:
(i) $d \mid a$ and $d \mid b$; that is, d is a common divisor of a and b;
(ii) $x \mid d$ for all $x \in \mathbb{N}$ such that $x \mid a$ and $x \mid b$; that is, every common divisor of a and b is also a divisor of d.

Remark 4.2. We note that the greatest common divisor of a and b is uniquely defined. Indeed, let d_1, d_2 be greatest common divisors of a and b. Using Definition 4.1 twice, we see that

$$\begin{aligned} d_1 \mid d_2, &\quad \text{that is, } \exists\, c_1 \in \mathbb{N} \colon d_2 = d_1 \cdot c_1; \\ d_2 \mid d_1, &\quad \text{that is, } \exists\, c_2 \in \mathbb{N} \colon d_1 = d_2 \cdot c_2. \end{aligned}$$

Substituting the first equality into the second yields

$$d_1 = d_1 \cdot c_1 \cdot c_2 \quad \Longleftrightarrow \quad 1 = c_1 \cdot c_2.$$

Remark 2.3 tells us that $c_1 = c_2 = 1$, whence $d_1 = d_2$, as asserted.

This uniqueness allows us to speak of *the* greatest common divisor of two natural numbers a and b. We denote this greatest common divisor by (a, b) and note that another notation in common use is $\gcd(a, b)$.

Theorem 4.3. *Let a, b be two natural numbers, not both equal to 0, with the prime factorizations*

$$a = \prod_{p \in \mathbb{P}} p^{a_p}, \quad b = \prod_{p \in \mathbb{P}} p^{b_p}.$$

The greatest common divisor (a, b) of a and b can be calculated as

$$(a, b) = \prod_{p \in \mathbb{P}} p^{d_p},$$

where $d_p := \min(a_p, b_p)$.

Proof. We set

$$d := \prod_{p \in \mathbb{P}} p^{d_p}.$$

Since the exponents $d_p = \min(a_p, b_p)$ are equal to 0 for all but a finite number of $p \in \mathbb{P}$, the natural number d is well defined. We have now to verify properties (i) and (ii) of Definition 4.1.

From the inequalities

$$d_p \le a_p \quad \text{and} \quad d_p \le b_p \quad \text{for all } p \in \mathbb{P},$$

we obtain at once from Lemma 3.5 (divisibility criterion) that

$$d \mid a \quad \text{and} \quad d \mid b.$$

Therefore, d is indeed a common divisor of a and b, and so property (i) is satisfied.

To verify property (ii) for d, let us choose an arbitrary common divisor x of a and b with prime decomposition

$$x = \prod_{p \in \mathbb{P}} p^{x_p}.$$

Again with the help of the divisibility criterion, we conclude that

$$x_p \le a_p, \quad x_p \le b_p,$$

for all prime numbers p, and we have, therefore,

$$x_p \le \min(a_p, b_p) = d_p.$$

A further application of the divisibility criterion yields $x \mid d$. Hence d also satisfies property (ii), and we have $d = (a,b)$. □

Example 4.4. Consider the natural numbers $a = 12 = 2^2 \cdot 3^1$ and $b = 15 = 3^1 \cdot 5^1$. Then the greatest common divisor (a,b) of a and b is given by

$$(a,b) = 2^0 \cdot 3^1 \cdot 5^0 = 3.$$

Remark 4.5. In the special case $a = 0$ and $b = 0$, we set $(a,b) := 0$.

Exercise 4.6. Determine $(3\,600, 3\,240)$, $\left(360^{360}, 540^{180}\right)$, and $\left(2^{32} - 1, 3^8 - 2^8\right)$.

Definition 4.7. Let a,b be nonzero natural numbers. A natural number m is called the *least common multiple* of a and b if it satisfies the following two properties:

(i) $a \mid m$ and $b \mid m$, that is, m is a common multiple of a and b;

(ii) $m \mid y$ for all $y \in \mathbb{N}$ with $a \mid y$ and $b \mid y$; that is, every common multiple of a and b is a multiple of m.

Remark 4.8. Analogously to the observation that we made in connection with the definition of the greatest common divisor, we can convince ourselves that the least common multiple is also a well-defined natural number. We denote it by $[a,b]$ and note that the notation $\operatorname{lcm}(a,b)$ is also used.

Theorem 4.9. *Let a,b be nonzero natural numbers with prime factorizations*

$$a = \prod_{p \in \mathbb{P}} p^{a_p}, \quad b = \prod_{p \in \mathbb{P}} p^{b_p}.$$

Then the least common multiple $[a,b]$ *of* a *and* b *is given by*

$$[a,b] = \prod_{p \in \mathbb{P}} p^{m_p},$$

where $m_p := \max(a_p, b_p)$.

Proof. We set

$$m := \prod_{p \in \mathbb{P}} p^{m_p}.$$

As in the proof of Theorem 4.3, we see that the natural number m is well defined. We have now to verify properties (i) and (ii) of Definition 4.7.

From the inequalities

$$m_p \geq a_p \quad \text{and} \quad m_p \geq b_p \quad \text{for all } p \in \mathbb{P},$$

it follows at once from Lemma 3.5 (divisibility criterion) that

$$a \mid m \quad \text{and} \quad b \mid m.$$

Therefore, m is in fact a common multiple of a and b, and so property (i) is satisfied.

To verify property (ii) for m, we choose an arbitrary common multiple y of a and b with prime factorization

$$y = \prod_{p \in \mathbb{P}} p^{y_p}.$$

Again using the divisibility criterion, we conclude that for all primes p, we have

$$y_p \geq a_p, \quad y_p \geq b_p,$$

and hence

$$y_p \geq \max(a_p, b_p) = m_p.$$

Using again the divisibility criterion, we see that $m \mid y$. Therefore, m satisfies property (ii) as well, and we have $m = [a,b]$. □

Remark 4.10. In the special case $a = 0$ or $b = 0$, we set $[a,b] := 0$.

Example 4.11. We again consider the example $a = 12 = 2^2 \cdot 3^1$ and $b = 15 = 3^1 \cdot 5^1$. Then for the least common multiple $[a,b]$ of a and b, we have

$$[a,b] = 2^2 \cdot 3^1 \cdot 5^1 = 60.$$

Remark 4.12. The notions of greatest common divisor and least common multiple can be extended inductively to more than two arguments. For n natural numbers $a_1,\dots,a_n$, the greatest common divisor $(a_1,\dots,a_n)$ is defined inductively as follows:

$$(a_1,\dots,a_n) := \big((a_1,\dots,a_{n-1}),a_n\big).$$

Analogously, the least common multiple $[a_1,\dots,a_n]$ of the natural numbers $a_1,\dots,a_n$ is defined inductively by

$$[a_1,\dots,a_n] := \big[[a_1,\dots,a_{n-1}],a_n\big].$$

In both cases, we may convince ourselves that we obtain the same result regardless of the order in which the numbers are arranged.

Exercise 4.13. Determine $(2880,3000,3240)$ and $[36,42,49]$.

Definition 4.14. We now define the notion of relative primality:
(i) Two natural numbers a,b are said to be *relatively prime* if their only common divisor is 1.
(ii) The natural numbers $a_1,\dots,a_n$ are said to be *relatively prime* if their only common divisor is 1.
(iii) The natural numbers $a_1,\dots,a_n$ are said to be *pairwise relatively prime* if each pair of numbers are relatively prime.

Exercise 4.15. Find three natural numbers a_1, a_2, a_3 that are relatively prime but not pairwise relatively prime.

Lemma 4.16. *We have the following facts about relative primality:*
(i) *For natural numbers a,b, we have $(a,b)\cdot[a,b] = a\cdot b$.*
(ii) *If the natural numbers $a_1,\dots,a_n$ are relatively prime, then $(a_1,\dots,a_n) = 1$.*
(iii) *If $a_1,\dots,a_n$ are pairwise relatively prime, then $[a_1,\dots,a_n] = a_1\cdots a_n$.*

Proof. (i) If a or b is equal to 0, the result is immediate. Otherwise, we consider the prime factorizations

$$a = \prod_{p\in\mathbb{P}} p^{a_p}, \quad b = \prod_{p\in\mathbb{P}} p^{b_p},$$

and set

$$d_p := \min(a_p,b_p), \quad m_p := \max(a_p,b_p).$$

By Theorems 4.3 and 4.9, we obtain

$$(a,b)\cdot[a,b] = \prod_{p\in\mathbb{P}} p^{d_p}\cdot\prod_{p\in\mathbb{P}} p^{m_p} = \prod_{p\in\mathbb{P}} p^{d_p+m_p} = \prod_{p\in\mathbb{P}} p^{a_p+b_p} = a\cdot b.$$

(ii) Our proof is by induction on n. The basis of the induction for $n = 2$ is established by the fact that for two relatively prime natural numbers a_1, a_2, we clearly have $(a_1, a_2) = 1$. We now assume that the assertion holds for $n \geq 2$ relatively prime natural numbers $a_1, \ldots, a_n$ and prove that on that assumption, the assertion holds for $n + 1$ relatively prime natural numbers. We distinguish two cases: $d := (a_1, \ldots, a_n)$ is equal to 1 and $d := (a_1, \ldots, a_n)$ is greater than 1. In the first case, we have immediately that

$$(a_1, \ldots, a_n, a_{n+1}) = (d, a_{n+1}) = (1, a_{n+1}) = 1.$$

In the second case, we obtain from the relative primality of $a_1, \ldots, a_{n+1}$ that d and a_{n+1} can have no common divisor greater than 1, from which it follows that

$$(a_1, \ldots, a_n, a_{n+1}) = (d, a_{n+1}) = 1.$$

This completes the proof of (ii) by induction.

(iii) We again carry out a proof by induction. The basis of the induction is the same as in part (ii). For the induction hypothesis, we now assume that the assertion holds for $n \geq 2$ pairwise relatively prime natural numbers $a_1, \ldots, a_n$, and we shall prove that on that assumption, the assertion holds for $n + 1$ pairwise relatively prime natural numbers. We begin by noting that

$$[a_1, \ldots, a_n, a_{n+1}] = \big[[a_1, \ldots, a_n], a_{n+1}\big] = [a_1 \cdots a_n, a_{n+1}].$$

Since $a_1, \ldots, a_n, a_{n+1}$ are pairwise relatively prime by assumption, we have in particular that $a_1 \cdots a_n$ and a_{n+1} are relatively prime. Thus the induction step follows from (i) and (ii), namely

$$\begin{aligned}[a_1, \ldots, a_n, a_{n+1}] &= 1 \cdot [a_1 \cdots a_n, a_{n+1}] \\ &= (a_1 \cdots a_n, a_{n+1}) \cdot [a_1 \cdots a_n, a_{n+1}] \\ &= a_1 \cdots a_n \cdot a_{n+1}.\end{aligned}$$

This completes the proof of part (iii). □

Exercise 4.17. Determine the conditions on the natural numbers $a_1, \ldots, a_n$ under which the following generalization of Lemma 4.16 is valid:

$$(a_1, \ldots, a_n) \cdot [a_1, \ldots, a_n] = a_1 \cdots a_n.$$

5. Division with Remainder

Let a, b be natural numbers. We assume for the moment that $b < a$. We consider the multiples $1 \cdot b, 2 \cdot b, 3 \cdot b, \ldots$, of b. It is clear that in a finite number of steps, we will arrive at a multiple of b that is strictly greater than a, at which point the previous multiple will be less than or equal to a. In mathematical

terms, this means that there exist natural numbers q,r such that

$$a = q \cdot b + r$$

with $0 \leq r < b$. We call this process and result *division of a by b with remainder r*. If $r = 0$, then b is a divisor of a.

Fig. 1. Division with remainder.

We shall now state and prove this obvious result.

Theorem 5.1 (Division with remainder). *Let a,b with $b \neq 0$ be natural numbers. Then there exist two uniquely determined natural numbers q,r with $0 \leq r < b$ such that*

$$a = q \cdot b + r. \tag{4}$$

Proof. We must show both the existence and uniqueness of the natural numbers q,r.

Existence: For every natural number q such that $q \cdot b \leq a$, we construct the natural number $r(q) := a - q \cdot b$. We then consider the set

$$\mathcal{M}(a,b) := \{r(q) \mid q \in \mathbb{N},\, q \cdot b \leq a\}.$$

With the choice $q = 0$, we establish that $r(0) = a$. Therefore, the set $\mathcal{M}(a,b)$ is not empty. By Lemma 1.21 (the well-ordering principle), there exists a natural number q_0 such that $r_0 := r(q_0)$ is the smallest element of $\mathcal{M}(a,b)$. The element r_0 satisfies the equality $r_0 = a - q_0 \cdot b$ if and only if

$$a = q_0 \cdot b + r_0. \tag{5}$$

We now show that $0 \leq r_0 < b$, that is, that (5) is the desired representation. Let us assume the contrary, namely that $r_0 \geq b$. Then there must exist $r_1 \in \mathbb{N}$ with $r_0 = b + r_1$. We note that $r_1 < r_0$, since $b \neq 0$ by assumption, whence $b > 0$. From the equivalent equations

$$b + r_1 = r_0 = a - q_0 \cdot b \quad \Longleftrightarrow \quad r_1 = a - (q_0 + 1) \cdot b,$$

it follows that $r_1 \in \mathcal{M}(a,b)$. We have already shown that $r_1 < r_0$, which contradicts the minimal choice of r_0. This completes the proof of the existence of the representation (4).

Uniqueness: Let q_1,r_1 and q_2,r_2 be natural numbers with $0 \leq r_1 < b$ and $0 \leq r_2 < b$ that satisfy the equations

$$a = q_1 \cdot b + r_1, \tag{6}$$
$$a = q_2 \cdot b + r_2. \tag{7}$$

Without loss of generality, we may assume that $r_2 \geq r_1$. From the inequalities that r_1 and r_2 satisfy, we obtain

$$0 \leq r_2 - r_1 < b.$$

Subtracting (6) from (7) yields the equality of natural numbers

$$r_2 - r_1 = (q_1 - q_2) \cdot b.$$

If we had $q_1 \neq q_2$, then we would have $q_1 - q_2 \geq 1$, whence

$$r_2 - r_1 = (q_1 - q_2) \cdot b \geq b.$$

But this contradicts the inequality $r_2 - r_1 < b$. We must therefore have $q_1 = q_2$, which at once yields $r_1 = r_2$. This completes the proof of the uniqueness of (4). □

Exercise 5.2. Carry out division with remainder for the following pairs of natural numbers: 773 and 337, $2^5 \cdot 3^4 \cdot 5^2$ and $2^3 \cdot 3^2 \cdot 5^3$, $2^{32} - 1$ and $4^8 + 1$.

Remark 5.3. Division with remainder is the basis of the *decimal representation* of natural numbers.

If $n \in \mathbb{N}$, $n \neq 0$, then there exists a maximal $\ell \in \mathbb{N}$ such that

$$n = q_\ell \cdot 10^\ell + r_\ell$$

with uniquely determined natural numbers $1 \leq q_\ell \leq 9$ and $0 \leq r_\ell < 10^\ell$. Treating the "remainder" r_ℓ in the same manner, one eventually arrives at the representation

$$n = q_\ell \cdot 10^\ell + q_{\ell-1} \cdot 10^{\ell-1} + \cdots + q_1 \cdot 10^1 + q_0 \cdot 10^0$$

with natural numbers $0 \leq q_j \leq 9$ ($j = 0, \ldots, \ell$) and $q_\ell \neq 0$. This leads to the decimal representation of the natural number n as the sequence of digits

$$n = q_\ell q_{\ell-1} \ldots q_1 q_0.$$

Exercise 5.4. Can this procedure be carried out for natural numbers $g > 0$ other than 10?

A. Prime Numbers: Facts and Conjectures

In this closing section, we present some interesting recent developments regarding one of the topics of this chapter, namely the prime numbers. We will exhibit a selection of deep results and unsolved conjectures.

A.1 Formulas for Prime Numbers

By Euclid's theorem, Theorem 2.10, we know that there are infinitely many prime numbers. It would be nice if we could find a "formula" for primes. The Mersenne primes and Fermat primes introduced in Example 2.14 offer such a formula only to a small extent. On the one hand, not all prime numbers are included, since, for example, the primes 11 and 13 are neither Mersenne primes nor Fermat primes. On the other hand, it remains an open question whether there are infinitely many Mersenne or Fermat primes.

An astounding result achieved in the 1970s by the Russian mathematician Yuri Matiyasevich is that there exists a polynomial of several variables with integer coefficients that generates all the prime numbers [6]. However, the polynomial was not given explicitly. Later, James Jones, Daihachiro Sato, Hideo Wada, und Douglas Wiens constructed the following polynomial in the 26 variables $A, B, \dots, Y, Z$,

$$\begin{aligned}
\mathcal{P}(A,B,\dots,Y,Z) &:= (K+2)\\
&\times \big(1-[WZ+H+J-Q]^2-[(GK+2G+K+1)\cdot(H+J)+H-Z]^2\\
&\quad -[16(K+1)^3(K+2)(N+1)^2+1-F^2]^2-[2N+P+Q+Z-E]^2\\
&\quad -[E^3(E+2)(A+1)^2+1-O^2]^2-[(A^2-1)Y^2+1-X^2]^2\\
&\quad -[16R^2Y^4(A^2-1)+1-U^2]^2-[N+L+V-Y]^2\\
&\quad -[(A^2-1)L^2+1-M^2]^2-[AI+K+1-L-I]^2\\
&\quad -[((A+U^2(U^2-A))^2-1)(N+4DY)^2+1-(X+CU)^2]^2\\
&\quad -[P+L(A-N-1)+B(2AN+2A-N^2-2N-2)-M]^2\\
&\quad -[Q+Y(A-P-1)+S(2AP+2A-P^2-2P-2)-X]^2\\
&\quad -[Z+PL(A-P)+T(2AP-P^2-1)-PM]^2\big),
\end{aligned}$$

and proved the following theorem.

Theorem A.1 (Jones, Sato, Wada, Wiens [4]). *For every prime number p, there exist natural numbers $a, b, \dots, y, z$ such that*

$$p = \mathcal{P}(a,b,\dots,y,z).$$

□

From an epistemological point of view, this formula is quite interesting. From a practical standpoint, however, it is not of immediate usefulness. If, for example, we are searching for very large primes, then Mersenne primes turn out to be very useful. The currently largest known prime number (as of 2016; see `http://primes.utm.edu/largest.html`) is

$$p = 2^{74\,207\,281} - 1.$$

It is a Mersenne prime with 22 338 618 digits. It begins thus:

$$\begin{aligned}
&30037641808460618205298609835916605005687586303 0\\
&30148484394169334554772321906799429689365530077 2\\
&68832044821488239942672783529070090483643221801 5\\
&34819965224137228768431021338628457366636150666 7\\
&53212277285935986405778025687564779586583214205 1\\
&17110963584426293657265038724071014798263132043 7\\
&14312911219839218876128850395877192035501718643 8\ldots
\end{aligned}$$

To print the entire number in a normal font size would take about 6 200 sheets of typewriter paper.

We might add that the current state of research in this area leaves much room for improvement, since there is to date no efficient closed formula for generating prime numbers.

A.2 Distribution of Primes

Since there does not seem to be a way of capturing the essence of prime numbers through formulas, one can ask instead for the probability that a randomly chosen natural number will turn out to be prime. To this end, we define the prime-counting function.

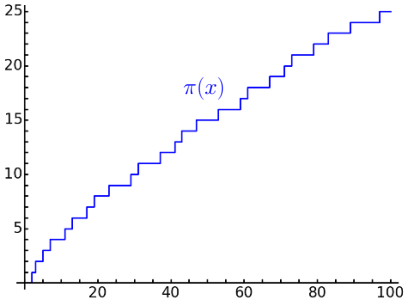
Definition A.2. For a positive real number x (the real numbers will be developed systematically in Chapter IV), the *prime-counting function* $\pi(x)$ is defined by

$$\pi(x) := \#\{p \in \mathbb{P} \mid p \leq x\};$$

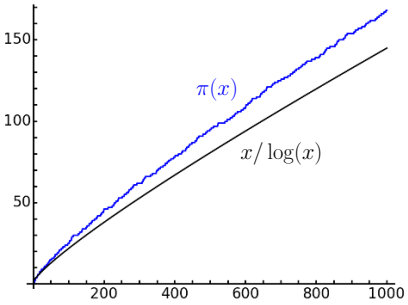
that is, $\pi(x)$ gives the number of primes less than or equal to x.

The probability that a randomly selected natural number in the interval $[0, x]$ is prime is given by the quotient $\pi(x)/x$.

The function $\pi(x)$ is a step function: whenever a new prime number appears, the value of the function increases by 1. In the interval $0 < x \leq 100$, one can easily determine that there are 25 primes.



Although at first glance, this function appears to reflect an irregularity in the prime numbers, when viewed on a larger scale, the function behaves quite regularly. For example, in the interval $0 < x \leq 1000$, we get the following graph:



This impression of regularity is confirmed by the prime number theorem, which was conjectured in a rudimentary form by Gauss, though it did not see a rigorous proof until the end of the nineteenth century, when the French mathematician Jacques Hadamard and the Belgian mathematician Charles-Jean de la Vallée Poussin gave independent proofs.

Theorem A.3 (Prime number theorem [2, 9]). *As $x \to \infty$, we have the asymptotic relation*

$$\pi(x) \sim \frac{x}{\log(x)};$$

that is, we have

$$\lim_{x\to\infty} \left(\frac{\pi(x)}{x/\log(x)} \right) = 1.$$

□

This result can also be formulated as follows: we have

$$\pi(x) = \frac{x}{\log(x)} + R(x),$$

with a "remainder term" $R(x)$ that as $x \to \infty$, grows more slowly than the function $x/\log(x)$. The question then naturally arises how large the growth of $R(x)$ can be.

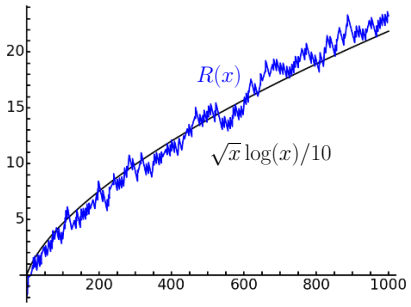
Conjecture (Remainder-term conjecture). *As $x \to \infty$, the remainder term $R(x)$ satisfies the estimate*

$$R(x) = O\left(\sqrt{x}\log(x)\right);$$

that is, there exists a positive constant C such that as $x \to \infty$, we have

$$|R(x)| \leq C\sqrt{x}\log(x).$$

For example, in the interval $0 < x \leq 1000$, we have the following graph:



It seems at present that mathematicians are a long way from being able to prove this estimate. The best estimates that one can prove right now are of the form

$$R(x) = O\left(x \cdot \exp\left(-D\log(x)^{3/5}\right)\right)$$

as $x \to \infty$ for some positive constant D, which go back to ideas of Ivan Vinogradov from the year 1958 [10].

A.3 Prime Gaps and Twin Primes

A prime gap describes the distance between two successive prime numbers. By the prime number theorem, we know that there are approximately $x/\log(x)$ prime numbers less than x. Therefore, the average gap between prime numbers less than x is about $\log(x)$. This observation must be considered in the light of the following two extreme cases.

The first is that it can be shown that there are arbitrarily large prime gaps. To see this, let $k \in \mathbb{N}$. We shall show that there exists a prime gap of length at least k. Let q denote the product of all prime numbers less than or equal to $k+1$. Then none of the k numbers

$$q+2, \ \ldots, \ q+k+1$$

is prime. We have thereby constructed a prime gap of length at least k.

The other extreme case has to do with the smallest possible gap between odd prime numbers, namely a sole even number lying between two primes. Such pairs are called *twin primes*. Here are a few simple examples of twin primes:

$$(5,7), \quad (11,13), \quad (17,19), \quad (29,31).$$

Conjecture (Twin prime conjecture). *There are infinitely many twin primes.*

This conjecture remains unproved. The largest known twin prime pair (as of 2016; see `http://primes.utm.edu/largest.html#twin`) is

$$\left(3756801695685 \cdot 2^{666669} - 1, \quad 3756801695685 \cdot 2^{666669} + 1\right).$$

We should not fail to mention here that in 2013, the Chinese-born American mathematician Yitang Zhang achieved a breakthrough toward a proof of the twin prime conjecture [12]. His result, as subsequently refined by other mathematicians, is that there are infinitely many pairs of prime numbers with a gap of length at most 272.

A.4 Riemann's Zeta Function

A study of Riemann's zeta function (after Georg Friedrich Bernhard Riemann) will lead us to a formulation of the Riemann hypothesis, which is equivalent to the remainder term conjecture. For this section, we refer the reader to Riemann's original work [8] as well as to Harold Edwards's book [1].

Definition A.4. For a real number $s > 1$, the *Riemann zeta function* is defined by the series

$$\zeta(s) := 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

For $s > 1$, the function $\zeta(s)$ is infinitely differentiable. The domain of definition of $\zeta(s)$ can be extended to the field $\mathbb{C}$ of complex numbers (we shall study the complex numbers systematically in Chapter V), and it turns out

that for $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$, $\zeta(s)$ is a holomorphic function (that is, it is complex differentiable at every point of the indicated domain).

At $s = 1$, the function becomes the harmonic series, which is well known to diverge, since for natural numbers N, one has, as $N \to \infty$, the relation

$$\sum_{n=1}^{N} \frac{1}{n} = \log(N) + \gamma + O\left(\frac{1}{N}\right),$$

where γ is the Euler–Mascheroni constant, whose value is $0.5772156649\ldots$.

For $\mathrm{Re}(s) > 1$, the Riemann zeta function can also be written as an infinite product. It is called the *Euler Product expansion* of $\zeta(s)$, and it is given by

$$\zeta(s) = \prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-s}}.$$

The validity of this product representation can be seen relatively easily as follows. Since $\mathrm{Re}(s) > 1$, we have for every prime $p \in \mathbb{P}$, the formula

$$\frac{1}{1 - p^{-s}} = \sum_{m=0}^{\infty} \frac{1}{p^{ms}} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots,$$

obtained with the help of geometric series, which leads to

$$\prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-s}} = \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \cdots\right)\left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \cdots\right)\cdots.$$

Multiplying out formally yields

$$\prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-s}} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{(2 \cdot 2)^s} + \frac{1}{5^s} + \frac{1}{(2 \cdot 3)^s} + \cdots.$$

In the numerators of the fractions on the right-hand side, every possible product of prime powers appears exactly once. By the fundamental theorem of arithmetic, Theorem 3.1, we know that each of these products is equal to a positive natural number, which proves the asserted equality. It can even be proved that the fundamental theorem of arithmetic is equivalent to the validity of the Euler product development.

Moreover, we see from the divergence of the harmonic series and the existence of the Euler product development as s approaches 1 from the right that there must be infinitely many prime numbers. That is, we have obtained an alternative proof of Theorem 2.10 of Euclid. In sum, we see that the Riemann zeta function somehow encodes fundamental arithmetic properties of the natural numbers.

To formulate the Riemann hypothesis, we begin by proving the theorem that the Riemann zeta function $\zeta(s)$ can be defined for an *arbitrary* complex argument s. To illustrate this deep result, we shall go into the proof rather extensively.

Theorem A.5. *The Riemann zeta function $\zeta(s)$ has a meromorphic continuation to the entire complex plane $\mathbb{C}$. It is holomorphic for all $s \in \mathbb{C}$ except for $s = 1$, where it has a simple pole. Moreover, for all $s \in \mathbb{C}$, the zeta function satisfies the following functional equation:*

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s).$$

Here $\Gamma(s)$ is Euler's gamma function, which is defined for $\operatorname{Re}(s) > 0$ by the formula

$$\Gamma(s) := \int_0^\infty e^{-x}\, x^{s-1}\, \mathrm{d}x.$$

Proof. We shall sketch the proof of this fundamental theorem. Using the Poisson summation formula, we obtain for $t > 0$ the equality

$$\sum_{k=-\infty}^{\infty} e^{-\pi k^2 t} = \frac{1}{\sqrt{t}} \sum_{k=-\infty}^{\infty} e^{-\pi k^2/t}.$$

If we define the function $\Theta(t)$ by

$$\Theta(t) := \sum_{k=1}^{\infty} e^{-\pi k^2 t},$$

we obtain the identity

$$2\Theta(t) + 1 = \frac{1}{\sqrt{t}}\left(2\Theta\left(\frac{1}{t}\right) + 1\right),$$

and by rearranging terms,

$$\Theta(t) = \frac{1}{\sqrt{t}}\left(\Theta\left(\frac{1}{t}\right) + \frac{1}{2}\right) - \frac{1}{2}. \tag{8}$$

If we replace s with $s/2$ in the definition of the gamma function and make the substitution $x \mapsto \pi n^2 x$, we obtain

$$\Gamma\left(\frac{s}{2}\right) = \pi^{s/2} n^s \int_0^\infty e^{-\pi n^2 x}\, x^{s/2-1}\, \mathrm{d}x.$$

If we solve this equation for $1/n^s$ and sum over all positive natural numbers, we obtain

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{\pi^{s/2}}{\Gamma(s/2)} \sum_{n=1}^{\infty} \int_0^{\infty} e^{-\pi n^2 x} \, x^{s/2-1} \, \mathrm{d}x.$$

The rapid rate of decay of the exponential function implies the uniform convergence of the integral on the right-hand side for $n \in \mathbb{N}$, $n > 1$. That allows us to invert the order of summation and integration. Using the definition of the function $\Theta(x)$, we obtain

$$\zeta(s) = \frac{\pi^{s/2}}{\Gamma(s/2)} \int_0^{\infty} \Theta(x) \, x^{s/2-1} \, \mathrm{d}x,$$

or

$$\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \int_0^1 x^{s/2-1} \, \Theta(x) \, \mathrm{d}x + \int_1^{\infty} x^{s/2-1} \, \Theta(x) \, \mathrm{d}x.$$

Application of (8) yields

$$\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \int_0^1 x^{s/2-1} \left(\frac{1}{\sqrt{x}} \Theta\left(\frac{1}{x}\right) + \frac{1}{2\sqrt{x}} - \frac{1}{2} \right) \mathrm{d}x + \int_1^{\infty} x^{s/2-1} \, \Theta(x) \, \mathrm{d}x.$$

The substitution $x \mapsto 1/x$ and a short calculation give us

$$\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \frac{1}{s(s-1)} + \int_1^{\infty} \left(x^{-s/2-1/2} + x^{s/2-1} \right) \Theta(x) \, \mathrm{d}x. \qquad (9)$$

The integral on the right-hand side converges for all $s \in \mathbb{C}$, since $\Theta(x)$ decays exponentially as $x \to \infty$, while $x^{-s/2-1/2} + x^{s/2-1}$ exhibits at most polynomial growth as $x \to \infty$. The first summand on the right-hand side has simple poles at $s = 0$ and at $s = 1$. But since $\Gamma(s)$ has a simple pole at $s = 0$, $\zeta(0)$ is well defined. We have obtained the desired meromorphic continuation of $\zeta(s)$ to the entire complex plane $\mathbb{C}$ with $\zeta(s)$ holomorphic everywhere except at $s = 1$, where it has a simple pole.

Using (9), we can see that the Riemann zeta function has a symmetry with respect to the transformation $s \mapsto 1 - s$, for we have

$$\begin{aligned}
&\pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s)\\
&\quad=\frac{1}{(1-s)((1-s)-1)}+\int_1^\infty\left(x^{-(1-s)/2-1/2}+x^{(1-s)/2-1}\right)\Theta(x)\,\mathrm{d}x\\
&\quad=\frac{1}{s(s-1)}+\int_1^\infty\left(x^{-s/2-1/2}+x^{s/2-1}\right)\Theta(x)\,\mathrm{d}x.
\end{aligned}$$

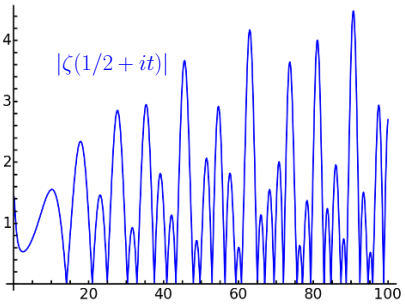
With (9), we obtain the asserted functional equation

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s)=\pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s),$$

which ends our proof sketch. □

With the help of (9) and the fact that the gamma function has poles at $s=0,-1,-2,\ldots$, we see at once that $\zeta(s)$ has zeros at $s=-2,-4,-6,\ldots$, called the *trivial zeros* of the Riemann zeta function. The Riemann hypothesis concerns the further zeros of $\zeta(s)$, which we discuss next. From the functional equation of the Riemann zeta function, we obtain, using our knowledge of the gamma function, that the behavior of $\zeta(s)$ for $\mathrm{Re}(s)<0$ can be derived from the known behavior of $\zeta(s)$ for $\mathrm{Re}(s)>1$. What is left is the behavior of $\zeta(s)$ in the strip $0\le\mathrm{Re}(s)\le 1$, which is therefore known as the *critical strip*. The line defined by the equation $\mathrm{Re}(s)=1/2$ plays a special role under the transformation $s\mapsto 1-s$, since it remains invariant.

Conjecture (Riemann hypothesis). *Aside from the trivial zeros of the Riemann zeta function $\zeta(s)$, namely those located at $s=-2,-4,-6,\ldots$, all the remaining zeros, the* nontrivial zeros, *lie on the* critical line $\mathrm{Re}(s)=1/2$.



About 10 trillion nontrivial zeros of the Riemann zeta function have been calculated numerically, and all of them lie on the critical line $\mathrm{Re}(s)=1/2$. In fact, it is known that infinitely many zeros of the Riemann zeta function

lie on the critical line. For example, the locations of the zeros of the function $|\zeta(1/2+it)|$ in the range $0 \leq t \leq 100$ are shown in the preceding graph.

The Riemann hypothesis remains to this day an open problem, and it is one of the most important, if not the most important, conjectures in number theory. It is one of the six remaining unsolved Millennium Problems. Its proof would have a host of implications. In particular, the vast number of theorems that have been proved conditionally on the validity of the Riemann hypothesis would at once become unconditional proofs.

A weak form of the Riemann hypothesis is given by the Lindelöf hypothesis (after Ernst Leonard Lindelöf), which also has defied all attempts at proof.

Conjecture (Lindelöf hypothesis [5]). *For every $\varepsilon > 0$, we have the following estimate for $t \gg 1$:*

$$\zeta\left(\frac{1}{2}+it\right) = o\left(t^{\varepsilon}\right).$$

Ralf Backlund a doctoral student of Ernst Lindelöf, showed that the Lindelöf hypothesis is equivalent to the statment

$$\#\left\{s \in \mathbb{C} \mid \zeta(s) = 0,\ \mathrm{Re}(s) \geq 1/2+\varepsilon,\ T \leq \mathrm{Im}(s) \leq T+1\right\} = o(\log(T)).$$

All that is known as of today is that the last quantity above is bounded by $O(\log(T))$ for $T \gg 1$.

A.5 The Goldbach Conjecture

By the fundamental theorem of arithmetic, Theorem 3.1, the prime numbers are the multiplicative building blocks of the natural numbers. If we now bring into play the other fundamental operation on the natural numbers, namely addition, there arise numerous questions that are not always easy to answer. One of the most famous problems of this sort goes back to a 1742 exchange of letters between Christian Goldbach and Leonhard Euler, where the question was raised whether it is possible to express every natural number greater than 5 as a sum of three primes. The following equivalent form is known as the Goldbach conjecture.

Conjecture (Goldbach conjecture). *Every even number greater than* 2 *can be written as the sum of two primes.*

This conjecture remains to this day only a conjecture, having withstood numerous attempts to prove it. Nonetheless, some progress has been made in the course of trying to prove the full conjecture. It is known, for example,

that every natural number greater than 6 can be written as the sum of distinct primes. A breakthrough was obtained by the Peruvian mathematician Harald Helfgott, who in 2013 announced a proof of the following weaker conjecture, known as the ternary Goldbach conjecture [3].

Conjecture (Ternary Goldbach conjecture). *Every odd number greater than* 5 *can be written as the sum of three primes.*

For his proof, Helfgott used the fact that it was already known that the ternary Goldbach conjecture is true for all odd numbers greater than $2 \cdot 10^{1346}$. Then, with the help of computers, he was able to verify the conjecture for odd numbers less than 10^{27}. The proof was then reduced to establishing the conjecture for odd numbers in the interval between those two numbers. Helfgott was able to do this using methods of analytic number theory, namely a refinement of the circle method, which goes back to Godfrey Harold Hardy, John Littlewood, and Ivan Vinogradov.

With this, we complete our brief tour of questions surrounding the prime numbers. For more on primes, we refer the interested reader to the vast literature on the subject. A good introduction can be found in the book [7], by Paulo Ribenboim, and the review article [11] by Don Zagier.

References

[1] H. M. Edwards: *Riemann's zeta function.* Dover, New York, 2001.

[2] J. Hadamard: *Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques.* Bull. Soc. Math. de France **24** (1896), 199–220.

[3] H. A. Helfgott: *The ternary Goldbach conjecture is true.* Preprint, December 30, 2013. Available online at `arXiv:1312.7748`.

[4] J. Jones, D. Sato, H. Wada, D. Wiens: *Diophantine representation of the set of prime numbers.* Amer. Math. Monthly **83** (1976), 449–464.

[5] E. Lindelöf: *Le calcul des résidus et ses applications dans la théorie des fonctions.* Gauthier-Villars, Paris, 1905.

[6] Y. Matiyasevich: *A Diophantine representation of the set of prime numbers.* Dokl. Akad. Nauk. SSSR **196** (1971), 770–773. English translation by R. N. Goss in Soviet Math. Dokl. **12** (1971), 249–254.

[7] P. Ribenboim: *The little book of bigger primes.* Springer, Berlin Heidelberg New York, 2004.

[8] B. Riemann: *Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse.* Monatsberichte der Königlichen Preußischen Akademie der Wissenschaften zu Berlin aus dem Jahre 1859 (1860), 671–680. In: Gesammelte Werke, Teubner, Leipzig, 1892. English translation by David R. Wilkins available online at `www.claymath.org/sites/default/files/ezeta.pdf`.

[9] C.-J. de la Vallée Poussin: *Recherches analytiques de la théorie des nombres premiers.* Ann. Soc. Scient. Bruxelles **20** (1896), 183–256.

[10] I. M. Vinogradov: *A new estimate of the function $\zeta(1+it)$.* Izv. Akad. Nauk. SSSR Ser. Mat. **22** (1958), 161–164.

[11] D. Zagier: *The first 50 million prime numbers.* Math. Intel. **0** (1977), 221–224.
[12] Y. Zhang: *Bounded gaps between primes.* Ann. of Math. (2) **179** (2014), 1121–1174.

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

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0_2

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

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, $\big(\mathrm{map}(A), \circ\big)$ is a semigroup.

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

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

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

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

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

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 $(\mathrm{map}(A_1), \circ)$ is an abelian semigroup but $(\mathrm{map}(A_2), \circ)$ 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

$$e \circ e' = e' = e' \circ e. \tag{3}$$

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

$$e' \circ e = e = e \circ e'. \tag{4}$$

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

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

This proves the uniqueness of the identity element. □

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 $\big(\mathrm{map}(A), \circ\big)$ is the identity mapping $\mathrm{id}_A : A \longrightarrow A$, 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:

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

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:

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

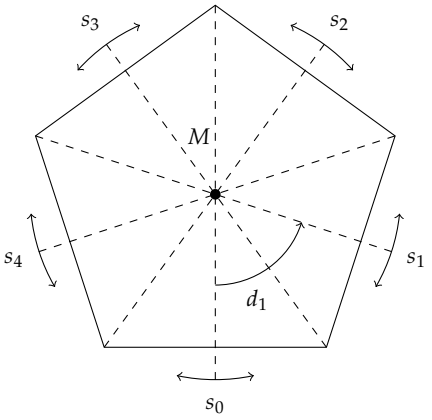


Fig. 1. Isometries of the regular pentagon.

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 *nth symmetric group*:

$$S_n = \big\{ \pi \,\big|\, \pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\} \text{ and } \pi \text{ is bijective} \big\}.$$

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 $\tau_j := \pi\big(\sigma(j)\big)$ for $1 \leq j \leq n$. The identity element is the identity permutation, given by the identity mapping on the set $\{1, \ldots, n\}$. Furthermore, the existence of the inverse of a permutation is guaranteed by the fact that every bijective mapping $\pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ 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.
(a) 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.
(b) Verify in detail the assertions of Example 2.8 (iii) regarding the dihedral group $(D_{2n}, \circ)$.
(c) 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

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

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

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

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

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

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 $\mathrm{ord}_G(g)$. If there is no such $n \in \mathbb{N}$, then the order of g is said to be infinite, that is, $\mathrm{ord}_G(g) := \infty$.

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

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

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

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 $\mathrm{ord}_G(g) = n$.

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 mth symmetric group S_m is a subgroup of the nth symmetric group S_n if we identify a permutation in S_m with the corresponding permutation of S_n that leaves $m+1, \ldots, n$ fixed. That is, $S_m \leq S_n$.

Exercise 2.23. Show that the rotations $\{d_0, \ldots, d_{n-1}\}$ 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*

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

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

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

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. □

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

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

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 $j \in \{0, \ldots, n-1\}$ and $k \in \{0, 1\}$. Show that the mapping $\mathrm{sgn} : (D_{2n}, \circ) \longrightarrow (\mathcal{R}_2, \oplus)$, 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* $\mathrm{im}(f) = H$.

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

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

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

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

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. □

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

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

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 $\mathrm{ord}_G(g) \geq \mathrm{ord}_H\big(f(g)\big)$.

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* $\mathrm{im}(f)$ *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:

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

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

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

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. □

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 $M_{m_1} \cap M_{m_2} \neq \varnothing$, where $\varnothing$ is the standard notation for the empty set. We must show that $M_{m_1} = M_{m_2}$. Since $M_{m_1} \cap M_{m_2} \neq \varnothing$, 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

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

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 $g_j \circ U$ ($j = 1, \ldots, k$) 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

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

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,

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

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

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

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

$$\pi_4 \circ A_3 = \{\pi_4 \circ \pi_1, \pi_4 \circ \pi_2, \pi_4 \circ \pi_3\} = \{\pi_4, \pi_5, \pi_6\},$$
$$A_3 \circ \pi_4 = \{\pi_1 \circ \pi_4, \pi_2 \circ \pi_4, \pi_3 \circ \pi_4\} = \{\pi_4, \pi_6, \pi_5\},$$

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

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

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

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:

$$(g_1 \circ N) \bullet (g_2 \circ N) := (g_1 \circ g_2) \circ N \qquad (g_1, g_2 \in G). \tag{7}$$

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 $\circ$,

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

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,

$$(g_1' \circ g_2') \circ N = \big(g_1 \circ (g_2 \circ h_1')\big) \circ N = (g_1 \circ g_2) \circ N;$$

here we have again used the associativity of $\circ$ and the equality $h_1' \circ N = N$. This completes the proof of the lemma. □

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

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

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. □

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

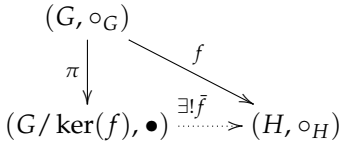
$$\begin{aligned}\pi(g_1 \circ_G g_2) &= (g_1 \circ_G g_2) \circ_G \ker(f) = \big(g_1 \circ_G \ker(f)\big) \bullet \big(g_2 \circ_G \ker(f)\big) \\ &= \pi(g_1) \bullet \pi(g_2);\end{aligned}$$

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*

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

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



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

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

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 $\bar{f}\big(g \circ \ker(f)\big) = f(g)$ $(g \in G)$. Let

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

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

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

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. □

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*

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

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

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 isomormphism

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

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*

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

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

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

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

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

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

$$h^* + x = h^* + y \quad \Longrightarrow \quad x = y$$

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 $(\mathrm{map}(A), \circ)$.
(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 $\circ_H$):

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

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:

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

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

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

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,

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

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

$$(h \circ a) \circ b = (h \circ b) \circ a \quad \Longleftrightarrow \quad \big((h \circ a), (h \circ b)\big) \sim (a, b).$$

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 $f : H \longrightarrow G$ satisfying

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

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 $f : H \longrightarrow G$ 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. □

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

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

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

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

(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,

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

that is, the pairs of natural numbers $(1, 3)$, $(2, 4)$, $(3, 5)$,... 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

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

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

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

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.

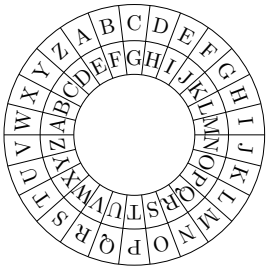
The history of cryptography goes back at least to the second century B.C., encrypted texts having been found as inscriptions on tombstones from that period. We are not, however, going to delve into the history of the subject, for which we refer the reader to the relevant literature, some of which is 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

$$a \equiv b \mod m \quad \Longleftrightarrow \quad m \mid (b - a)$$

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

Remark B.2. If $a \equiv c \mod m$ and $b \equiv d \mod m$, then we have the following two rules:
(i) $a + b \equiv c + d \mod m$,
(ii) $a \cdot b \equiv c \cdot d \mod m$.
Rule (ii) shows in particular that if $a \equiv c \mod m$, then for every $n \in \mathbb{N}$, we have the congruence

$$a^n \equiv c^n \mod m.$$

Example B.3. Let $m = 22$. Then, for example, we have $23 \equiv 1 \mod 22$, $47 \equiv 3 \mod 22$, and $87 \equiv 21 \mod 22$. Using the above rules of calculation, we obtain $23 + 47 \equiv 1 + 3 \equiv 4 \mod 22$ and $47 \cdot 87 \equiv 3 \cdot 21 \equiv 19 \mod 22$, as well as $47^{17} \equiv 3^{17} \equiv 129\,140\,163 \equiv 9 \mod 22$.

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

$$a \equiv b \mod m \quad \Longleftrightarrow \quad R_m(a) = R_m(b).$$

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 $\overline{a}$ or $a \bmod m$. The residue classes modulo m are given by the set

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

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*

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

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 $x' \equiv x \mod m$.*

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

$$a \cdot x \equiv 1 \mod m,$$

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

$$a \cdot x \equiv 1 \equiv a \cdot x' \mod m \quad \Longleftrightarrow \quad a(x - x') \equiv 0 \mod m.$$

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

$$x' \equiv x \mod m.$$

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

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 $43 \cdot x \equiv 1 \mod 88464$ is solved by

$$x \equiv -20573 \mod 88464.$$

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

$$\texttt{xgcd}(43,88\,464) = (1,-20573,10),$$

which means that $(43,88\,464) = 1$ and that we have the equality

$$43 \cdot (-20\,573) + 88\,464 \cdot 10 = 1.$$

Finally, if we calculate

$$\texttt{mod}(-20\,573,88\,464) = 67\,891,$$

we obtain a solution x such that $0 < x < 88\,464$.

Remark B.9. A solution x of the congruence $a \cdot x \equiv 1 \mod m$ with $0 < x < m$ can also be obtained with the SAGE command a`.inverse_mod`(m). For example, for $a = 43$ and $m = 88\,464$, we obtain the result

$$43\,\texttt{.inverse\_mod}(88\,464) = 67\,891.$$

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*

$$a^p \equiv a \mod p.$$

Proof. If a if a multiple of p, then we have $a \equiv 0 \mod p$ and therefore also $a^p \equiv 0 \mod p$, 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

$$a \cdot j \equiv j' \mod p.$$

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

$$a \cdot j_1 \equiv j' \equiv a \cdot j_2 \mod p \quad \Longleftrightarrow \quad p \mid a(j_1 - j_2)$$

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

$$(a \cdot 1) \cdots (a \cdot (p-1)) \equiv 1 \cdots (p-1) \mod p,$$

that is,

$$a^{p-1}(p-1)! \equiv (p-1)! \mod p \quad \Longleftrightarrow \quad (a^{p-1}-1)(p-1)! \equiv 0 \mod p.$$

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

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

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

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 $a^m \mod m$, obtaining

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

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*

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

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 $a \equiv 0 \mod p \cdot q$, and by the rules for calculating with congruences,

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

which proves the asserted congruence.

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

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

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 $b^{q-1} \equiv 1 \mod q$, that is,

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

On multiplication by a, we obtain the congruence

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

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

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

(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

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

Analogously, one proves the further congruence

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

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

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

This completes the proof. □

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $c \equiv a \mod m$ has been validated. We can see this as follows: We have the congruences $c \equiv b^x \mod m$ and $b \equiv a^k \mod m$, and therefore,

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

Here the integer x is uniquely determined by the conditions $k \cdot x \equiv 1 \mod n$ 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

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

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

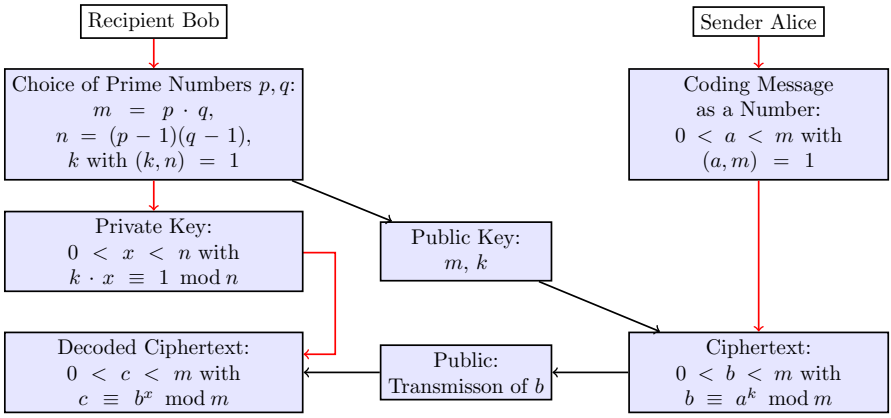
$$c \equiv a \cdot 1^y \equiv a \mod m,$$

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 $a^k \mod m$, is simple. The same holds for calculating $b^x \mod m$, 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

$$m = p \cdot q = 229 \cdot 389 = 89\,081,$$
$$n = (p-1) \cdot (q-1) = 228 \cdot 388 = 88\,464.$$

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

$$m = 89\,081 \quad \text{and} \quad k = 43.$$

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

$$a = 8\,073.$$

She then calculates the integer b with $b \equiv 8\,073^{43} \mod 89\,081$ and $0 < b < 89\,081$, and transmits the encrypted message

$$b = 30\,783.$$

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

$$x = 67\,891.$$

Bob can now decrypt the ciphertext $b = 30\,783$ by calculating the uniquely determined integer c such that $c \equiv 30\,783^{67\,891} \mod 89\,081$ and $0 < c < 89\,081$. He obtains

$$c = 8\,073,$$

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

$$p = 207472224677348520782169522210760858748099647472111729275299258991219668475054965831008441673255007 7,$$

$$q = 721261014729547490954452378504349240996938214818676546008250008539351955652592145558870542302075142 1.$$

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 $b \equiv a^k \mod m$ and $0 < b < m$ and transmits the encrypted message

$$\begin{aligned} b = &36057960785874251250398847578656552489665279269698 \\ &48103809148617096444525775586803496118061034800457 \\ &72014608577306857911068935474951466578892598687245 \\ &60731528213013240247453503444393031326009131733 84. \end{aligned}$$

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 $k \cdot x \equiv 1 \mod n$ 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

$$x = 104401135324870505502450484338388986847546073886829960755895271025580076987630920517547436138532178250756334845499692617255424236824270979867936841467996764131302675920269540039352109260472016033311107.$$

Bob now decodes the message b using the SAGE command

$$\texttt{power\_mod}(b, x, m)$$

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

$$c = 808273776932788577666982833265826932858369708576 33,$$

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. Cilberto, 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.

III The Rational Numbers

1. The Integers and Divisibility Theory

In the last section of Chapter II, we introduced the (additive) group of integers $(\mathbb{Z}, +)$, which we constructed with the help of Theorem 6.5 of that chapter from the (additive) semigroup $(\mathbb{N}, +)$ of natural numbers. We now recall that the natural numbers also have the structure of a monoid with respect to the multiplication defined in Chapter I. Our first task is to generalize this multiplicative structure to the set of integers. To this end, we return to the definition of $\mathbb{Z}$ as a set of equivalence classes (see the proof of Theorem 6.5 of Chapter II); that is,

$$\mathbb{Z} = \{[a, b] \mid (a, b) \in \mathbb{N} \times \mathbb{N}\}.$$

We now define the *product* of two integers $[a, b]$ and $[a', b']$ by the formula

$$[a, b] \cdot [a', b'] := [aa' + bb', ab' + a'b]. \tag{1}$$

Here, as in Chapter I, we write $aa' + bb'$ and $ab' + a'b$ as an abbreviated form for the natural numbers $(a \cdot a') + (b \cdot b')$ and $(a \cdot b') + (a' \cdot b)$. To establish that this multiplication operation is well defined, we must prove that the product (1) is independent of the choice of representatives (a, b) and (a', b'). To this end, let (c, d) and (c', d') be representatives of the equivalence classes $[a, b]$ and $[a', b']$; that is, we have

$$a + d = b + c \quad \text{and} \quad a' + d' = b' + c'. \tag{2}$$

We must demonstrate the equivalence class equality

$$[aa' + bb', ab' + a'b] = [cc' + dd', cd' + c'd].$$

We therefore define the natural number

$$n := (a' + b')(c + d) = a'c + a'd + b'c + b'd.$$

We then calculate, keeping in mind the equalities (2),

$$\begin{aligned} aa' + bb' + cd' + c'd + n &= a'(a + d) + b'(b + c) + c(a' + d') + d(b' + c') \\ &= a'(b + c) + b'(a + d) + c(b' + c') + d(a' + d') \\ &= ab' + a'b + cc' + dd' + n. \end{aligned}$$

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0_3

Since $\mathbb{N}$ is regular, we obtain, after canceling n, the equivalence

$$(aa' + bb', ab' + a'b) \sim (cc' + dd', cd' + c'd),$$

from which follows the asserted equality of the equivalence classes.

With the notation $a - b = [a, b]$, the definition (1) takes the familiar form

$$(a - b) \cdot (a' - b') = (aa' + bb') - (ab' + a'b),$$

and we obtain at once the sign rules (for $m, n \in \mathbb{N}$)

$$m \cdot (-n) = -(m \cdot n) = (-m) \cdot n, \quad (-m) \cdot (-n) = m \cdot n.$$

For ease of notation, we shall hereinafter write $-m \cdot n$ instead of $-(m \cdot n)$. As with the multiplication of natural numbers, we shall frequently suppress the dot indicating multiplication. We leave the proof of the following lemma to the reader.

Lemma 1.1. *The operation of multiplication defined by* (1) *on the set of integers is associative and commutative. That is, for all integers a, b, c, we have*

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c \quad \textit{and} \quad a \cdot b = b \cdot a.$$

Furthermore, for all integers a, b, c, we have the distributive laws

$$(a + b) \cdot c = a \cdot c + b \cdot c \quad \textit{and} \quad a \cdot (b + c) = a \cdot b + a \cdot c.$$

□

Exercise 1.2. Prove the rules for multiplication of integers given in Lemma 1.1.

Putting everything together, we have the following.

Remark 1.3. The set of integers $\mathbb{Z}$ has defined on it two operations, addition $+$ and multiplication $\cdot$. We express this by writing $(\mathbb{Z}, +, \cdot)$. Both operations satisfy the associative and commutative laws. Addition and multiplication are linked by the two distributive laws. The set $(\mathbb{Z}, +)$ with addition is an abelian group with identity element 0; the inverse element to $a \in \mathbb{Z}$ is denoted by $-a$. The set $(\mathbb{Z}, \cdot)$ with multiplication is an abelian monoid with identity element 1. The elements other than ± 1 have no multiplicative inverse. That is, the integers ± 1 are the only elements of $\mathbb{Z}$ possessing a multiplicative inverse in $\mathbb{Z}$.

We shall now extend the divisibility theory that we developed for the natural numbers in Section 2 of Chapter I to the integers. In analogy to Definition 2.1 of Chapter I, we note that the integer $b \neq 0$ *divides* the integer a if

there exists an integer c such that $a = b \cdot c$. The notion of a *common divisor* of two integers carries over directly from Definition 2.1 of Chapter I. The validity of the divisibility rules from Lemma 2.4 of Chapter I also carries over directly to the integers. By a simple generalization of Remark 2.6 of Chapter I, we call the divisors $1, -1, a, -a$, or $\pm 1, \pm a$ for short, the *trivial divisors* of the integer a. Furthermore, we call two integers a, b *associates* if they differ by at most a sign, that is, if $a = \pm b$. A *prime number* p is now characterized as an integer greater than 1 that has only the trivial divisors ± 1 and $\pm p$. Lemma 2.9 of Chapter I carries over directly to the integers.

In general, one can carry out division with remainder on integers, just as with the natural numbers.

Theorem 1.4 (Division with remainder, revisited). *Let a, b be integers with $b \neq 0$. Then there exist uniquely determined integers q, r with $0 \leq r < |b|$ such that*

$$a = q \cdot b + r. \tag{3}$$

Proof. The proof is a simple modification of the proof of Theorem 5.1 of Chapter I and is left to the reader as an exercise. □

Exercise 1.5. Carry out the proof of Theorem 1.4.

A significant difference between the divisibility theory for the natural numbers and that for the integers is the fact that in the integers, it is possible to invoke Euclid's lemma as an aid in proving the fundamental theorem of arithmetic, in contrast to how we proceeded in the case of the natural numbers, where Euclid's lemma appeared only as a consequence of the fundamental theorem. In this direction, we prove the following lemma.

Lemma 1.6. *Let a, b be relatively prime integers, that is, such that a, b have only the trivial divisors ± 1 in common. Then there exist integers x, y such that*

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

Proof. We consider the set of all integer linear combinations of a and b, that is, the set

$$\mathfrak{a} := \{x_1 \cdot a + y_1 \cdot b \mid x_1, y_1 \in \mathbb{Z}\} \subseteq \mathbb{Z}.$$

Since we have either $a \in \mathfrak{a} \cap \mathbb{N}$ or $-a \in \mathfrak{a} \cap \mathbb{N}$, the intersection $\mathfrak{a} \cap \mathbb{N}$ is nonempty. By the well-ordering principle (Lemma 1.21 of Chapter I), there exists a least positive element $d \in \mathfrak{a} \cap \mathbb{N}$. We must show that in fact, $d = 1$.

We note first that since $d \in \mathfrak{a}$, there exist integers x_0 and y_0 such that $d = x_0 \cdot a + y_0 \cdot b$. Now let $c \in \mathfrak{a}$ be an arbitrary element of the form $c = x_1 \cdot a + y_1 \cdot b$ with $x_1, y_1 \in \mathbb{Z}$. On dividing c by d with remainder, we obtain $q, r \in \mathbb{Z}$, $0 \leq r < d$, so that we have

$$c = q \cdot d + r. \tag{4}$$

If we now substitute $c = x_1 \cdot a + y_1 \cdot b$ and $d = x_0 \cdot a + y_0 \cdot b$ in (4), we obtain

$$x_1 \cdot a + y_1 \cdot b = q(x_0 \cdot a + y_0 \cdot b) + r,$$

which is equivalent to

$$r = (x_1 - q \cdot x_0)a + (y_1 - q \cdot y_0)b \in \mathfrak{a} \cap \mathbb{N}.$$

If we had $r \neq 0$, then we would also have $0 < r < d$. But that would contradict the minimality of the choice of $d \in \mathfrak{a} \cap \mathbb{N}$. We therefore have $r = 0$, and so $d \mid c$. From the representation $c = x_1 \cdot a + y_1 \cdot b$ with the special values $x_1 = 1$, $y_1 = 0$ and $x_1 = 0$, $y_1 = 1$, we have $d \mid a$ and $d \mid b$; that is, d is a common divisor of a and b. But the integers a, b have only the trivial common divisors ± 1, and so we must have $d = 1$. Finally, if we set $x := x_0$ and $y := y_0$, then we obtain

$$x \cdot a + y \cdot b = d = 1,$$

as asserted. □

Lemma 1.7 (Euclid's lemma, revisited). *Let a, b be integers and p a prime number. If $p \mid a \cdot b$, then we must have $p \mid a$ or $p \mid b$.*

Proof. We begin with the divisibility relationship $p \mid a \cdot b$. If $p \mid a$, then we are done. If, on the other hand, $p \nmid a$, then we must prove $p \mid b$. Since p is prime and $p \nmid a$, we conclude that a and p are relatively prime. By the previous lemma, there exist, therefore, integers x, y such that

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

On multiplying this equality by b, we obtain

$$b = x \cdot ab + yb \cdot p. \tag{5}$$

The divisibility rules of Lemma 2.4 of Chapter I (extended to the integers) now show us that p divides the right-hand side of (5), whence it also divides the left-hand side. That is, we have $p \mid b$, as asserted. □

For the set of integers, the fundamental theorem of arithmetic takes the following form:

Theorem 1.8 (Fundamental theorem of arithmetic, revisited). *Every nonzero integer a can be represented in the form*

$$a = e \cdot p_1^{a_1} \cdots p_r^{a_r},$$

which is the product of $e \in \{\pm 1\}$ and a product of r ($r \in \mathbb{N}$) powers of the distinct prime numbers $p_1, \ldots, p_r$ with positive natural-number exponents $a_1, \ldots, a_r$. This representation is unique up to the order of the factors.

Proof. For the absolute value $|a|$ of a, we have

$$a = e \cdot |a|$$

with uniquely determined sign $e \in \{\pm 1\}$. The existence and uniqueness of the prime decomposition of the natural number $|a|$ can be inferred from the proof of Theorem 3.1 of Chapter I.

As an alternative to the uniqueness proof there, one can in the present situation complete the proof quickly and elegantly using induction and Euclid's lemma. We leave this as an exercise for the reader. □

Exercise 1.9. Carry out the uniqueness proof of the fundamental theorem of arithmetic using Lemma 1.7.

If a, b are integers, then the definition of the *greatest common divisor* (a, b) of a and b can be reduced to Definition 4.1 of Chapter I of the greatest common divisor of natural numbers by setting

$$(a, b) := (|a|, |b|).$$

Likewise, the definition of the *least common multiple* $[a, b]$ of a and b can be reduced to Definition 4.7 of Chapter I of the least common multiple of natural numbers by setting

$$[a, b] := [|a|, |b|].$$

By carrying over the divisibility criterion given in Lemma 3.5 of Chapter I to the integers, we obtain at once the analogues of Theorems 4.3 and 4.9 of Chapter I for calculating the greatest common divisor and least common multiple of integers a and b using prime decompositions.

2. Rings and Subrings

The example presented in the previous section of the integers with the two operations addition and multiplication that are linked by the distributive laws is the prototype for the following definition of a ring.

Definition 2.1. A nonempty set R with two operations (generally called addition $+$ and multiplication $\cdot$) that satisfies the following properties is called a *ring*:

(i) $(R, +)$ is an abelian group.

(ii) $(R, \cdot)$ is a semigroup.

(iii) For all $a, b, c \in R$, we have the two distributive laws

$$(a + b) \cdot c = a \cdot c + b \cdot c,$$
$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Definition 2.2. A ring $(R, +, \cdot)$ is said to be *commutative* if for all $a, b \in R$, we have the equality $a \cdot b = b \cdot a$.

Remark 2.3. (i) We call the identity element of the additive group $(R, +)$ of a ring $(R, +, \cdot)$ the *zero element* and denote it by 0. We denote the additive inverse of $a \in R$ by $-a$. We define the *difference* of elements $a, b \in R$ by $a - b := a + (-b)$.

(ii) The ring $(R, +, \cdot)$ consisting solely of the zero element 0 is called the *zero ring* or *null ring* and is denoted by $(\{0\}, +, \cdot)$.

(iii) If the multiplicative semigroup $(R, \cdot)$ in a ring $(R, +, \cdot)$ that is not the zero ring is a monoid, we call its identity element the *unit element* and denote it by 1. The unit element is uniquely determined and necessarily satisfies the inequality $1 \neq 0$, since we have $R \neq \{0\}$.

(iv) To simplify notation, we will, as usual, agree that multiplication takes precedence over addition (and therefore, $a \cdot b + c$ means $(a \cdot b) + c$).

Example 2.4. (i) $(\mathbb{Z}, +, \cdot)$ is a commutative ring with unit element.

(ii) $(\mathcal{R}_n, \oplus, \odot)$ is a commutative ring with unit element.

(iii) $(2 \cdot \mathbb{Z}, +, \cdot)$ is a commutative ring, but it does not have a unit element, since $1 \notin 2 \cdot \mathbb{Z}$.

(iv) The following example is well known from linear algebra. We consider the set of 2×2 matrices with integer entries, that is, the set

$$\mathrm{M}_2(\mathbb{Z}) := \left\{ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\middle|\; a, b, c, d \in \mathbb{Z} \right\}.$$

Two matrices $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $A' = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$ are added and multiplied as follows:

$$A + A' = \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} := \begin{pmatrix} a + a' & b + b' \\ c + c' & d + d' \end{pmatrix}$$

and

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

where the addition and multiplication of the individual entries are the usual operations on integers. We leave it to the reader as an exercise to show that $(\mathrm{M}_2(\mathbb{Z}), +, \cdot)$ is a ring with unit element. The zero and unit elements of $\mathrm{M}_2(\mathbb{Z})$ are given by the matrices

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};$$

the additive inverse of the matrix A is

$$-A := \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}.$$

We note that this ring is not commutative.

(v) Let $(R, +, \cdot)$ be a ring. We define the *polynomial ring* $(R[X], +, \cdot)$ in the variable X with coefficients in R as the set

$$R[X] := \left\{ \sum_{j \in \mathbb{N}} a_j \cdot X^j \,\middle|\, a_j \in R, a_j = 0 \text{ for all but finitely many } j \in \mathbb{N} \right\}$$

with the operations

$$\left(\sum_{j \in \mathbb{N}} a_j \cdot X^j \right) + \left(\sum_{j \in \mathbb{N}} b_j \cdot X^j \right) := \sum_{j \in \mathbb{N}} (a_j + b_j) \cdot X^j,$$
$$\left(\sum_{j \in \mathbb{N}} a_j \cdot X^j \right) \cdot \left(\sum_{j \in \mathbb{N}} b_j \cdot X^j \right) := \sum_{j \in \mathbb{N}} \left(\sum_{\substack{k,\ell \in \mathbb{N} \\ k+\ell=j}} (a_k \cdot b_\ell) \right) \cdot X^j.$$

We leave it as an exercise to the reader to show that $(R[X], +, \cdot)$ is a ring. We remark that we have denoted the formal variable in the elements of a polynomial ring by the capital letter X to distinguish between the polynomial $p(X) \in R[X]$ and its value $p(x) \in R$ at the element $x \in R$.

Exercise 2.5. Prove that the polynomial ring $(R[X], +, \cdot)$ from Example 2.4 (v) is a ring and that it is commutative if and only if $(R, +, \cdot)$ is commutative.

Exercise 2.6. Let A be a nonempty set and $(R, +_R, \cdot_R)$ a ring. Prove that the set $\mathrm{map}(A, R)$ of all mappings from A to R is a ring when it is equipped with the two operations

$$(f, g) \mapsto f + g, \text{ with } (f + g)(a) := f(a) +_R g(a) \quad (f, g \in \mathrm{map}(A, R), a \in A),$$
$$(f, g) \mapsto f \cdot g, \text{ with } (f \cdot g)(a) := f(a) \cdot_R g(a) \quad (f, g \in \mathrm{map}(A, R), a \in A).$$

Exercise 2.7. Determine which of the ring properties from Definition 2.1 are satisfied by the set $\mathbb{N}$ with the operations "max" as the addition and $+$ as the multiplication.

Lemma 2.8. *Let $(R, +, \cdot)$ be a ring. Then for $a, b, c \in R$, we have the following:*
(i) $a \cdot 0 = 0 \cdot a = 0$.
(ii) $a \cdot (-b) = (-a) \cdot b = -a \cdot b$.
(iii) $(-a) \cdot (-b) = a \cdot b$.
(iv) $(a - b) \cdot c = a \cdot c - b \cdot c$.
(v) $a \cdot (b - c) = a \cdot b - a \cdot c$.

Proof. (i) By the distributive law, we have $a \cdot a = a \cdot (a + 0) = a \cdot a + a \cdot 0$, and it therefore follows by adding $-a \cdot a$ to both sides that $a \cdot 0 = 0$. The equality $0 \cdot a = 0$ follows analogously.

(ii) Using (i) and the distributive law, we obtain the equality

$$a \cdot b + a \cdot (-b) = a\big(b + (-b)\big) = a \cdot 0 = 0,$$

from which the asserted equality $a \cdot (-b) = -a \cdot b$ follows by addition of $-a \cdot b$ to both sides. The second equality, $(-a) \cdot b = -a \cdot b$, follows analogously.

(iii) Using (ii), we compute

$$(-a) \cdot (-b) = a \cdot \big(-(-b)\big) = a \cdot b.$$

(iv) Using the distributive law and (ii), we calculate

$$(a - b) \cdot c = \big(a + (-b)\big) \cdot c = a \cdot c + (-b) \cdot c = a \cdot c - b \cdot c.$$

(v) The proof of (v) is analogous to that of (iv). □

Definition 2.9. An element $a \neq 0$ of a ring $(R, +, \cdot)$ is called a *left zero divisor* if there exists $b \in R$, $b \neq 0$, such that $a \cdot b = 0$. One defines *right zero divisors* analogously. If the ring is commutative, we may speak simply of *zero divisors*.

A ring $(R, +, \cdot)$ is called a *domain* if it has no (left or right) zero divisors. A nonnull commutative ring $(R, +, \cdot)$ without zero divisors is called an *integral domain*.

Example 2.10. (i) The ring $(\mathbb{Z}, +, \cdot)$ is an integral domain.

(ii) The rings $(\mathcal{R}_n, \oplus, \odot)$ are in general not integral domains, since as a rule, they possess zero divisors. For example, the element $2 \in \mathcal{R}_6$ is a zero divisor, since we have $2 \odot 3 = 0$.

(iii) The noncommutative matrix ring $\mathrm{M}_2(\mathbb{Z})$ also has zero divisors. The matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, for example, is both a left and right zero divisor, since

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

Exercise 2.11. Generalize Example 2.10 (ii) as follows: if $n > 1$ is not prime, then $(\mathcal{R}_n, \oplus, \odot)$ has zero divisors.

Exercise 2.12. Show that if $(R, +, \cdot)$ is an integral domain, then so is the polynomial ring $(R[X], +, \cdot)$.

Exercise 2.13. Does the ring $(\mathrm{map}(A, R), +, \cdot)$ from Exercise 2.6 have zero divisors?

Lemma 2.14. *Let $(R, +, \cdot)$ be a domain with unit element 1. If there exists a positive natural number n such that*

$$n \cdot 1 := \underbrace{1 + \cdots + 1}_{n \text{ times}} = 0,$$

and n is the minimal positive number with this property, then n is prime.

Proof. We proceed with a proof by contradiction. Assume, then, that the number n is not prime. Then there exist natural numbers $k, \ell \in \mathbb{N}$ with $1 < k, \ell < n$ such that $n = k \cdot \ell$. We thereby obtain

$$n \cdot 1 = (k \cdot \ell) \cdot 1 = (k \cdot 1) \cdot (\ell \cdot 1) = 0.$$

Since $(R, +, \cdot)$ is without zero divisors, it follows that

$$k \cdot 1 = 0 \text{ or } \ell \cdot 1 = 0.$$

But this contradicts the minimality of n. □

Definition 2.15. Let $(R, +, \cdot)$ be a domain with unit element 1 and suppose that there exists a positive natural number p, which we take to be minimal, with the property $p \cdot 1 = 0$. Then by the above lemma, p must be prime, and we call p the *characteristic* of the ring R; we write $\text{char}(R) = p$.

If there is no positive natural number n such that $n \cdot 1 = 0$, we say that R has *characteristic zero*.

Example 2.16. (i) The ring of integers $(\mathbb{Z}, +, \cdot)$ has characteristic zero, since there is no positive natural number n such that $n \cdot 1 = 0$.

(ii) If p is a prime, then the ring $(\mathcal{R}_p, \oplus, \odot)$ is an integral domain. Its characteristic can easily be seen to be

$$\text{char}(\mathcal{R}_p) = p,$$

since for all $k \in \{1, \ldots, p-1\}$, we have $k \cdot 1 \neq 0$, but $p \cdot 1 = 0$ in $\mathcal{R}_p$.

Definition 2.17. Let $(R, +, \cdot)$ be a ring with unit element 1 and let $a \in R$ be an arbitrary element. An element $b \in R$ is called a *left inverse* of a if $b \cdot a = 1$. Similarly, an element $c \in R$ is a *right inverse* of a if $a \cdot c = 1$.

An element $d \in R$ is called a *(multiplicative) inverse of a* if it is both a left and right inverse, that is, if $a \cdot d = d \cdot a = 1$. If $a \in R$ has a multiplicative inverse, we denote it by a^{-1} or $\frac{1}{a}$ or sometimes $1/a$.

An element $a \in R$ that has a multiplicative inverse in R is called a *unit*.

Example 2.18. In the ring $(\mathbb{Z}, +, \cdot)$, the elements $a \neq \pm 1$ do not have a multiplicative inverse. The units of $(\mathbb{Z}, +, \cdot)$ are $+1$ and -1.

Exercise 2.19. What are the units of the polynomial ring $(\mathbb{Z}[X], +, \cdot)$?

Exercise 2.20. Show that the units of a ring $(R, +, \cdot)$ with unit element 1 form a group with respect to the ring operation of multiplication.

Exercise 2.21. Determine the group of units for each of the rings $(\mathcal{R}_n, \oplus, \odot)$, $n = 5, 8, 10, 12$. Which of these groups of units are isomorphic?

Definition 2.22. Let $(R, +, \cdot)$ be a ring. A subset $S \subseteq R$ is called a *subring* of R if the restriction of the operations $+, \cdot$ to S (which for simplicity we again denote by $+, \cdot$) define a ring structure on S, that is, if $(S, +, \cdot)$ is itself a ring. We express this relationship by writing $S \leq R$.

Lemma 2.23 (Subring criterion). *Let $(R, +, \cdot)$ be a ring, and $S \subseteq R$ a nonempty subset. Then we have the equivalence*

$$S \leq R \quad \Longleftrightarrow \quad a - b \in S, a \cdot b \in S \quad \forall a, b \in S.$$

Proof. (i) If S is a subring of R, then clearly the difference $a - b$ and the product $a \cdot b$ must be in S for all $a, b \in S$.

(ii) Conversely, suppose that $a - b \in S$, $a \cdot b \in S$ for all $a, b \in S$. Since S is nonempty, we have from the subgroup criterion, namely Lemma 2.25 of Chapter II, that $(S, +)$ is an abelian subgroup of the additive group $(R, +)$. Since we also have $a \cdot b \in S$ for all $a, b \in S$, it follows that S is closed under multiplication. Furthermore, S inherits the associative law with respect to multiplication and the distributive laws from R. Therefore, $(S, +, \cdot)$ is a ring, and thus the proof is complete. □

Example 2.24. The ring $(2\mathbb{Z}, +, \cdot)$ of even integers is a subring of the ring $(\mathbb{Z}, +, \cdot)$ of integers.

Exercise 2.25. Find additional examples of subrings of the ring $(\mathbb{Z}, +, \cdot)$ of integers.

Exercise 2.26. Let $(R, +, \cdot)$ be a ring. Is $(R, +, \cdot)$ a subring of the ring of polynomials $(R[X], +, \cdot)$?

3. Ring Homomorphisms, Ideals, and Quotient Rings

In the previous section, we defined for rings the subsidiary object of a subring in analogy to the definition of a subgroup of a group, introduced in Chapter II. In pursuit of further structural analysis of groups, we also introduced in Chapter II the notions of group homomorphism, normal subgroup, and quotient group. Continuing in that vein in our discussion of

rings, we shall now introduce analogous definitions adapted to the more complex structure of a ring. We begin with the concept of a ring homomorphism.

Definition 3.1. Let $(R, +_R, \cdot_R)$ and $(S, +_S, \cdot_S)$ be rings. A mapping

$$f : (R, +_R, \cdot_R) \longrightarrow (S, +_S, \cdot_S)$$

is called a *ring homomorphism* if for all $r_1, r_2 \in R$, we have the equalities

$$f(r_1 +_R r_2) = f(r_1) +_S f(r_2),$$
$$f(r_1 \cdot_R r_2) = f(r_1) \cdot_S f(r_2).$$

For two rings to be related by a ring homomorphism thus means that the images under f of the sum and product of r_1 and r_2 in R are equal to the corresponding sum and product of the images of r_1 and r_2 in S. One says that the mapping f *preserves the ring structure*.

A bijective (that is, injective and surjective) ring homomorphism is called a *ring isomorphism*. If $f : (R, +_R, \cdot_R) \longrightarrow (S, +_S, \cdot_S)$ is a ring isomorphism, we say that the rings R and S are *isomorphic* and write $R \cong S$.

Exercise 3.2. Determine which of the following mappings are ring homomorphisms. Let $(R, +, \cdot)$ be a nonnull ring, and A a nonempty set:

(a) $f_1 : R[X] \longrightarrow R$, where $f_1\left(\sum_{j\in\mathbb{N}} a_j \cdot X^j\right) := a_0$.

(b) $f_2 : R[X] \longrightarrow R$, where $f_2\left(\sum_{j\in\mathbb{N}} a_j \cdot X^j\right) := a_1$.

(c) $f_3 : \mathrm{map}(A, R) \longrightarrow R$, where $f_3(g) := r$ $(g \in \mathrm{map}(A, R))$ for a fixed $r \in R$.

(d) $f_4 : \mathrm{map}(A, R) \longrightarrow R$, where $f_4(g) := g(a)$ $(g \in \mathrm{map}(A, R))$ for a fixed $a \in A$.

(e) $f_5 : R[X] \longrightarrow R$, where $f_5\left(\sum_{j\in\mathbb{N}} a_j \cdot X^j\right) := \sum_{j\in\mathbb{N}} a_j \cdot r^j$ for a fixed $r \in R$.

In analogy to group homomorphisms introduced in Chapter II, we define the kernel and image of a ring homomorphism.

Definition 3.3. Let $(R, +_R, \cdot_R)$ be a ring with zero element 0_R, and let $(S, +_S, \cdot_S)$ be a ring with zero element 0_S. Furthermore, let $f : (R, +_R, \cdot_R) \longrightarrow (S, +_S, \cdot_S)$ be a ring homomorphism. Then

$$\ker(f) := \{r \in R \mid f(r) = 0_S\}$$

is called the *kernel* of f, and

$$\operatorname{im}(f) := \{s \in S \mid \exists r \in R : s = f(r)\}$$

is called the *image* of f.

Lemma 3.4. *Let $f : (R, +_R, \cdot_R) \longrightarrow (S, +_S, \cdot_S)$ be a ring homomorphism. Then* $\ker(f)$ *is a subring of R, and* $\operatorname{im}(f)$ *is a subring of S.*

Proof. The proof proceeds along the same lines as that of Lemma 3.10 of Chapter II and can therefore be left to the reader as an exercise. □

Exercise 3.5. Prove Lemma 3.4.

Exercise 3.6. Determine the kernel and image of those mappings in Exercise 3.2 that are ring homomorphisms.

Remark 3.7. For ease of notation, we shall usually omit the subscripts on the operations $+_R$ and $\cdot_R$ and on the zero element 0_R.

Example 3.8. We build here on Example 7.8 of Chapter II, in which we introduced the group homomorphism $f : (\mathbb{Z}, +) \longrightarrow (\mathcal{R}_n, \oplus)$ via the assignment $a \mapsto R_n(a)$. It is easily verified that this mapping induces a surjective ring homomorphism

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

For the kernel, we have, as in Example 7.8 of Chapter II,

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

Remark 3.9. The kernel $\ker(f)$ of a ring homomorphism $f : (R, +, \cdot) \longrightarrow (S, +, \cdot)$ is, by Lemma 3.4, a subring of R. We also note that the products $r \cdot a$ and $a \cdot r$ are in the kernel of f not only for all $a, r \in \ker(f)$, but for all $a \in \ker(f)$ and $r \in R$, since we have

$$\begin{aligned} f(r \cdot a) &= f(r) \cdot f(a) = f(r) \cdot 0 = 0, \\ f(a \cdot r) &= f(a) \cdot f(r) = 0 \cdot f(r) = 0. \end{aligned}$$

This observation leads to the following definition.

Definition 3.10. Let $(R, +, \cdot)$ be a ring. A subgroup $(\mathfrak{a}, +)$ of the additive group $(R, +)$ is called an *ideal* of R if the products

$$r \cdot a \quad \text{and} \quad a \cdot r$$

are in $\mathfrak{a}$ for all $a \in \mathfrak{a}$ and all $r \in R$, that is, if we have the inclusions

$$\begin{aligned} R \cdot \mathfrak{a} &:= \{r \cdot a \mid r \in R,\ a \in \mathfrak{a}\} \subseteq \mathfrak{a}, \\ \mathfrak{a} \cdot R &:= \{a \cdot r \mid r \in R,\ a \in \mathfrak{a}\} \subseteq \mathfrak{a}. \end{aligned}$$

Remark 3.11. An ideal $\mathfrak{a}$ of a ring $(R, +, \cdot)$ is automatically also a subring of R. The converse of this statement is, however, not in general true.

Example 3.12. (i) Let $(R, +, \cdot)$ be a ring. The subgroup $(\mathfrak{a}, +) = (\{0\}, +)$ is clearly an ideal of R. We call it the *zero ideal* of R and denote it by (0).

(ii) Again, let $(R, +, \cdot)$ be a ring. The additive group $(\mathfrak{a}, +) = (R, +)$ is an ideal of R. If R has a unit element 1, then this ideal is also called the *unit ideal* of R and is denoted by (1).

(iii) Let $(R, +, \cdot)$ be a commutative ring. For a fixed $a \in R$, we consider the set

$$\mathfrak{a} := \{a \cdot r \mid r \in R\}.$$

We can see that $\mathfrak{a}$ is an ideal of R. Since $0 \in \mathfrak{a}$, it follows that $\mathfrak{a}$ is not empty. If in addition, we have $a \cdot r_1, a \cdot r_2 \in \mathfrak{a}$, then we also have that the difference

$$a \cdot r_1 - a \cdot r_2 = a \cdot (r_1 - r_2)$$

is in $\mathfrak{a}$. By the subgroup criterion, Lemma 2.25 of Chapter II, it follows that $(\mathfrak{a}, +)$ is a subgroup of the additive group $(R, +)$. Finally, if we have $a \cdot r \in \mathfrak{a}$ and $s \in R$, then by the associativity and commutativity of multiplication, we have

$$s \cdot (a \cdot r) = a \cdot (r \cdot s) \in \mathfrak{a};$$

that is, we have $R \cdot \mathfrak{a} \subseteq \mathfrak{a}$, and by commutativity, $\mathfrak{a} \cdot R \subseteq \mathfrak{a}$. Therefore, $\mathfrak{a}$ is an ideal of R. We call it the *principal ideal* generated by a and denote it by (a).

Exercise 3.13. Let $(R, +, \cdot)$ be a ring with unit element 1, and let $\mathfrak{a} \subseteq R$ be an ideal of R with $1 \in \mathfrak{a}$. Show that we must have $\mathfrak{a} = R$.

Exercise 3.14. Is there a subring of $(\mathbb{Z}, +, \cdot)$ that is not an ideal of $\mathbb{Z}$?

Exercise 3.15. Find a subring of the polynomial ring $(\mathbb{Z}[X], +, \cdot)$ that is not an ideal of $\mathbb{Z}[X]$.

Exercise 3.16. Give examples of ideals in the polynomial ring $(\mathbb{Z}[X], +, \cdot)$. Are there any ideals in this ring that are not principal ideals?

Lemma 3.17. *Let $f : (R, +, \cdot) \longrightarrow (S, +, \cdot)$ be a ring homomorphism. Then* $\ker(f)$ *is an ideal of R.*

Proof. From Lemma 3.10 of Chapter II, we see that $\big(\ker(f), +\big)$ is an additive subgroup of $(R, +)$. From Remark 3.9, we derive the inclusions

$$R \cdot \ker(f) \subseteq \ker(f) \quad \text{and} \quad \ker(f) \cdot R \subseteq \ker(f),$$

which prove that $\ker(f)$ is an ideal as claimed. □

Exercise 3.18. Which of the kernels of the ring homomorphisms in Exercise 3.2 are principal ideals?

Lemma 3.19. *In the ring $(\mathbb{Z}, +, \cdot)$, all the ideals are principal, that is, for each ideal $\mathfrak{a}$, there exists an integer a such that $\mathfrak{a} = (a)$.*

Proof. If $\mathfrak{a}$ is the zero ideal, then we have $\mathfrak{a} = (0)$, and we are done. Otherwise, since $\mathfrak{a}$ is not the zero ideal, there exists a nonzero integer $b \in \mathfrak{a}$. Multiplying b by -1 if necessary, we have a nonzero element in the set $\mathfrak{a} \cap \mathbb{N}$. By the well-ordering principle, there exists a least positive integer $a \in \mathfrak{a}$.

By the definition of an ideal, we see at once that we must have

$$(a) \subseteq \mathfrak{a}.$$

We now prove the reverse inclusion. To this end, let $c \in \mathfrak{a}$ be an arbitrary element. On dividing c by a with remainder (see Theorem 1.4), we obtain uniquely determined integers q, r with $0 \leq r < a$ such that

$$c = q \cdot a + r.$$

Since we have $a, c \in \mathfrak{a}$, it follows from the fact that $\mathfrak{a}$ is an ideal that we must also have $r = c - q \cdot a$ as an element of $\mathfrak{a}$. If we had $r \neq 0$, then r would be a nonzero element of $\mathfrak{a} \cap \mathbb{N}$ that is smaller than a. But that would contradict the minimality of a. We must therefore have $r = 0$, and we have $c = q \cdot a$, that is, we have $c \in (a)$. This completes the proof of the inclusion $\mathfrak{a} \subseteq (a)$. We have thus completed the proof that $\mathfrak{a} = (a)$, which is what was to be shown. □

Definition 3.20. Let $(R, +, \cdot)$ be a ring, and $\mathfrak{a}$ an ideal of R. Since the additive group $(R, +)$ is abelian by definition, the additive subgroup $(\mathfrak{a}, +)$ of the ideal is automatically a normal subgroup of $(R, +)$. We may therefore consider the quotient group $(R/\mathfrak{a}, \oplus)$. The elements of $R/\mathfrak{a}$ are given by cosets of the form $r + \mathfrak{a}$ ($r \in R$). The sum of two cosets $r_1 + \mathfrak{a}$ and $r_2 + \mathfrak{a}$ is given by

$$(r_1 + \mathfrak{a}) \oplus (r_2 + \mathfrak{a}) = (r_1 + r_2) + \mathfrak{a}$$

(see Definition 5.1 of Chapter II). We note that in contrast to Definition 5.1 of Chapter II, where we denoted the operation in the quotient group by $\bullet$, here we have chosen the notation $\oplus$ to account for the additive nature of this construction. Incidentally, the reader will have no difficulty in distinguishing the notation $\oplus$ used here from the same notation used in the example of $(\mathcal{R}_n, \oplus, \odot)$.

We define a multiplicative operation $\odot$ on the quotient group $(R/\mathfrak{a}, \oplus)$ by setting, for two cosets $r_1 + \mathfrak{a}$ and $r_2 + \mathfrak{a}$ (in this case as well, there should be no confusion with $(\mathcal{R}_n, \oplus, \odot)$),

$$(r_1 + \mathfrak{a}) \odot (r_2 + \mathfrak{a}) := (r_1 \cdot r_2) + \mathfrak{a}. \tag{6}$$

This definition appears to depend on the choice of representatives r_1 and r_2 for the cosets $r_1 + \mathfrak{a}$ and $r_2 + \mathfrak{a}$. In the following theorem, we prove in particular that the multiplicative operation $\odot$ is, in fact, well defined.

Theorem 3.21. *Let $(R, +, \cdot)$ be a ring, and $\mathfrak{a}$ an ideal of R. Then the set of cosets $R/\mathfrak{a}$ together with the two operations*

$$(r_1 + \mathfrak{a}) \oplus (r_2 + \mathfrak{a}) = (r_1 + r_2) + \mathfrak{a},$$
$$(r_1 + \mathfrak{a}) \odot (r_2 + \mathfrak{a}) = (r_1 \cdot r_2) + \mathfrak{a},$$

forms a ring.

Proof. (i) First, we see from Definition 3.20 that $(R/\mathfrak{a}, \oplus)$ is an abelian group with identity element (zero element) $\mathfrak{a}$.

(ii) We now show that the multiplicative operation $\odot$ is well defined. To this end, let r_1, r_1' and r_2, r_2' be representatives of the respective cosets $r_1 + \mathfrak{a}$ and $r_2 + \mathfrak{a}$. To prove that multiplication $\odot$ is well defined, we must prove the equality

$$(r_1 \cdot r_2) + \mathfrak{a} = (r_1' \cdot r_2') + \mathfrak{a}. \tag{7}$$

We have the equalities

$$r_1' = r_1 + a_1 \quad (a_1 \in \mathfrak{a}),$$
$$r_2' = r_2 + a_2 \quad (a_2 \in \mathfrak{a}),$$

connecting the representatives r_1, r_1' and r_2, r_2'. From these, we calculate

$$r_1' \cdot r_2' = (r_1 + a_1) \cdot (r_2 + a_2) = r_1 \cdot r_2 + r_1 \cdot a_2 + a_1 \cdot r_2 + a_1 \cdot a_2.$$

Because $\mathfrak{a}$ is an ideal, we see that

$$r_1 \cdot a_2 + a_1 \cdot r_2 + a_1 \cdot a_2 \in \mathfrak{a}.$$

Therefore, the product $r_1' \cdot r_2'$ is also a representative of the coset $(r_1 \cdot r_2) + \mathfrak{a}$; that is, we have indeed the asserted equality (7). This proves that the multiplication $\odot$ is well defined.

(iii) The associativity of the multiplication $\odot$ is shown, using definition (6) and the associativity of the multiplication operation $\cdot$, as follows:

$$\begin{aligned}(r_1 + \mathfrak{a}) \odot \big((r_2 + \mathfrak{a}) \odot (r_3 + \mathfrak{a})\big) &= (r_1 + \mathfrak{a}) \odot \big((r_2 \cdot r_3) + \mathfrak{a}\big) \\ &= \big(r_1 \cdot (r_2 \cdot r_3)\big) + \mathfrak{a} = \big((r_1 \cdot r_2) \cdot r_3\big) + \mathfrak{a} \\ &= \big((r_1 \cdot r_2) + \mathfrak{a}\big) \odot (r_3 + \mathfrak{a}) = \big((r_1 + \mathfrak{a}) \odot (r_2 + \mathfrak{a})\big) \odot (r_3 + \mathfrak{a}).\end{aligned}$$

(iv) The proof of the distributive laws also derives from definition (6) and the distributive laws in the ring $(R, +, \cdot)$; for example, we have

$$\begin{aligned}(r_1+\mathfrak{a})\odot\big((r_2+\mathfrak{a})\oplus(r_3+\mathfrak{a})\big) &= (r_1+\mathfrak{a})\odot\big((r_2+r_3)+\mathfrak{a}\big)\\ &= \big(r_1\cdot(r_2+r_3)\big)+\mathfrak{a} = \big((r_1\cdot r_2)+(r_1\cdot r_3)\big)+\mathfrak{a}\\ &= \big((r_1\cdot r_2)+\mathfrak{a}\big)\oplus\big((r_1\cdot r_3)+\mathfrak{a}\big)\\ &= (r_1+\mathfrak{a})\odot(r_2+\mathfrak{a})\oplus(r_1+\mathfrak{a})\odot(r_3+\mathfrak{a}).\end{aligned}$$

We have proved that $(R/\mathfrak{a},\oplus,\odot)$ is a ring. □

Definition 3.22. Let $(R,+,\cdot)$ be a ring, and $\mathfrak{a}$ an ideal of R. Then the ring $(R/\mathfrak{a},\oplus,\odot)$ is called the *quotient ring* of R by the ideal $\mathfrak{a}$.

Remark 3.23. Let $f:(R,+,\cdot)\longrightarrow(S,+,\cdot)$ be a ring homomorphism. Then, Lemma 3.17 tells us that $\ker(f)$ is an ideal of R; by Theorem 3.21, we may consider the quotient ring $(R/\ker(f),\oplus,\odot)$. We recognize the canonical group homomorphism

$$\pi:(R,+)\longrightarrow(R/\ker(f),\oplus)$$

from Remark 5.6 of Chapter II, defined by the assignment $r\mapsto r+\ker(f)$, as a ring homomorphism, since we have

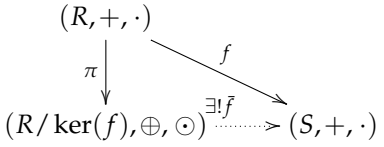
$$\begin{aligned}\pi(r_1\cdot r_2) &= (r_1\cdot r_2)+\ker(f)\\ &= \big(r_1+\ker(f)\big)\odot\big(r_2+\ker(f)\big)\\ &= \pi(r_1)\odot\pi(r_2).\end{aligned}$$

We call this the *canonical ring homomorphism*.

Theorem 3.24 (Homomorphism theorem for rings). *Let $f:(R,+,\cdot)\longrightarrow(S,+,\cdot)$ be a ring homomorphism. Then f induces a uniquely determined injective ring homomorphism*

$$\bar{f}:(R/\ker(f),\oplus,\odot)\longrightarrow(S,+,\cdot)$$

such that $\bar{f}\big(r+\ker(f)\big)=f(r)$ for all $r\in R$. This result can be illustrated schematically by saying that the following diagram is commutative:



Commutativity of the diagram means that the same result is obtained by executing the mapping f directly or by first executing π and then the mapping $\bar{f}$.

Proof. According the homomorphism theorem for groups, Theorem 5.7 of Chapter II, there exists a uniquely determined injective group homomor-

phism

$$\bar{f} : (R/\ker(f), \oplus) \longrightarrow (S, +)$$

such that $\bar{f}\big(r + \ker(f)\big) = f(r)$ for all $r \in R$. It thus remains to show that $\bar{f}$ respects the multiplicative structure as well. Using the definition of the operation $\odot$, the definition of $\bar{f}$, and the ring homomorphism f, we compute the image under $\bar{f}$ of the product of the two cosets $r_1 + \ker(f)$ and $r_2 + \ker(f)$ as

$$\begin{aligned}\bar{f}\big(\big(r_1 + \ker(f)\big) \odot \big(r_2 + \ker(f)\big)\big) &= \bar{f}\big((r_1 \cdot r_2) + \ker(f)\big) = f(r_1 \cdot r_2) \\ &= f(r_1) \cdot f(r_2) = \bar{f}\big(r_1 + \ker(f)\big) \cdot \bar{f}\big(r_2 + \ker(f)\big).\end{aligned}$$

We have thus shown that $\bar{f}$ is a ring homomorphism, which completes the proof of the homomorphism theorem for rings. □

Corollary 3.25. *Let $f : (R, +, \cdot) \longrightarrow (S, +, \cdot)$ be a surjective ring homomorphism. Then f determines a uniquely determined ring isomorphism*

$$\bar{f} : (R/\ker(f), \oplus, \odot) \cong (S, +, \cdot)$$

such that $\bar{f}\big(r + \ker(f)\big) = f(r)$ for all $r \in R$. □

Example 3.26. (i) We continue Example 3.8, in which we saw that there arises a surjective ring homomorphism

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

with $\ker(f) = n\mathbb{Z}$. By Corollary 3.25 to the homomorphism theorem for rings, we have the ring isomorphism

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

given by the assignment $a + n\mathbb{Z} \mapsto R_n(a)$.

(ii) Let $(R, +, \cdot) = (\mathbb{Z}, +, \cdot)$, and let $(S, +, \cdot)$ be a domain with unit element 1. The assignment

$$n \mapsto \begin{cases} n \cdot 1 = \underbrace{1 + \cdots + 1}_{n \text{ times}}, & \text{if } n \in \mathbb{Z}, n \geq 0, \\ -\big((-n) \cdot 1\big), & \text{if } n \in \mathbb{Z}, n < 0, \end{cases}$$

defines a ring homomorphism $f : (\mathbb{Z}, +, \cdot) \longrightarrow (S, +, \cdot)$. The kernel of f is equal to the ideal

$$\ker(f) = \{n \in \mathbb{Z} \mid n \cdot 1 = 0\}.$$

We distinguish two cases:

(a) $\operatorname{char}(S) = 0$: In this case, we have by definition that $n \cdot 1 \neq 0$ for all $n \in \mathbb{Z} \setminus \{0\}$; that is, $\ker(f) = \{0\}$, which implies the injectivity of f. Thus every ring of characteristic zero contains a subring isomorphic to the ring of integers $(\mathbb{Z}, +, \cdot)$; in the sequel, we will identify this subring with the integers $(\mathbb{Z}, +, \cdot)$.
(b) $\operatorname{char}(S) = p$: In this case, we have by definition that for the prime number p, we have $p \cdot 1 = 0$; that is, $\ker(f) = p\mathbb{Z}$. By the homomorphism theorem for rings, we thus obtain an injective ring homomorphism $\bar{f} : \mathbb{Z}/p\mathbb{Z} \longrightarrow S$. Therefore every ring of characteristic p contains (an isomorphic copy of) the quotient ring $(\mathbb{Z}/p\mathbb{Z}, \oplus, \odot) \cong (\mathcal{R}_p, \oplus, \odot)$ as a subring.

Exercise 3.27. Find a ring homomorphism $f : (\mathbb{Z}[X], +, \cdot) \longrightarrow (\mathbb{Z}, +, \cdot)$ such that for some $a \in \mathbb{Z}$, there exists, as described in Corollary 3.25, a ring isomorphism

$$(\mathbb{Z}[X]/(X - a), \oplus, \odot) \cong (\mathbb{Z}, +, \cdot).$$

Exercise 3.28. Formulate and prove an analogue for rings of the group isomorphism from Exercise 5.11 of Chapter II.

4. Fields and Skew Fields

The motivation to extend the definition of a ring to define fields and skew fields is again based on the desire to remove restrictions on the solutions to linear equations. If $(R, +, \cdot)$ is a commutative ring with unit element 1, then the equation

$$a \cdot x = b \quad (a, b \in R) \tag{8}$$

is solvable in R if a has an inverse in R, in which case the solution is $x = a^{-1} \cdot b$. A field is a commutative ring with unit element 1 such that each of its nonzero elements has a multiplicative inverse in R, with the result that (8) always has a solution in R except for the case $a = 0$ and $b \neq 0$.

Definition 4.1. Let $(R, +, \cdot)$ be a ring with unit element 1. Then we denote the set of units of R by $R^\times$; that is,

$$R^\times = \{a \in R \mid a \text{ has a (multiplicative) inverse in } R\}.$$

A ring $(R, +, \cdot)$ with unit element 1 is called a *skew field* if

$$R^\times = R \setminus \{0\}.$$

A commutative skew field is called a *field*.

Remark 4.2. (i) Let $(R,+,\cdot)$ be a ring with unit element 1. Then $(R^\times,\cdot)$ is a group with identity element 1. We call it the *multiplicative group of the ring* $(R,+,\cdot)$.

(ii) If $(R,+,\cdot)$ is a skew field, then every $a \in R$, $a \neq 0$, has a multiplicative inverse $a^{-1} = \frac{1}{a} = 1/a \in R$. The multiplicative group of the skew field $(R,+,\cdot)$ is equal to $(R \setminus \{0\},\cdot)$.

(iii) If $(R,+,\cdot)$ is a field and $a, b \in R$ with $b \neq 0$, then we use the notation

$$a \cdot b^{-1} = \frac{a}{b} = a/b.$$

Example 4.3. Let p be a prime number. Then the ring $(\mathcal{R}_p,\oplus,\odot)$ is a skew field; it is, in fact, a field. The situation is especially simple for the case $p = 2$, for which have the field comprising the two elements $0, 1$.

Exercise 4.4. Try to find a skew field with finitely many elements that is not a field.

Remark 4.5. In Chapter VI, we shall discuss an example of a skew field that is not a field, namely the *Hamiltonian quaternions*.

Lemma 4.6. *Let $(K,+,\cdot)$ be a field. Then for $a, b, c, d \in K$, we have the following calculational rules:*

(i) *If $b, c \neq 0$, then*

$$\frac{a}{b} = \frac{a \cdot c}{b \cdot c}.$$

(ii) *If $b, d \neq 0$, then*

$$\frac{a}{b} \pm \frac{c}{d} = \frac{a \cdot d \pm b \cdot c}{b \cdot d}.$$

(iii) *If $b, d \neq 0$, then*

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d}.$$

Proof. (i) For $b, c \neq 0$, we compute

$$\frac{a}{b} = a \cdot b^{-1} = a \cdot c \cdot c^{-1} \cdot b^{-1} = (a \cdot c) \cdot (b \cdot c)^{-1} = \frac{a \cdot c}{b \cdot c}.$$

(ii) Using the commutativity of multiplication and the distributive laws, we can calculate, for $b, d \neq 0$,

$$\begin{aligned} \frac{a}{b} \pm \frac{c}{d} &= a \cdot b^{-1} \pm c \cdot d^{-1} \\ &= (a \cdot d) \cdot (b \cdot d)^{-1} \pm (b \cdot c) \cdot (b \cdot d)^{-1} \\ &= (a \cdot d \pm b \cdot c) \cdot (b \cdot d)^{-1} \\ &= \frac{a \cdot d \pm b \cdot c}{b \cdot d}. \end{aligned}$$

(iii) Using the commutativity of multiplication, we can calculate, for $b, d \neq 0$,

$$\frac{a}{b} \cdot \frac{c}{d} = (a \cdot b^{-1}) \cdot (c \cdot d^{-1}) = (a \cdot c) \cdot (b \cdot d)^{-1} = \frac{a \cdot c}{b \cdot d}.$$

This completes the proof of the lemma. □

5. Construction of Fields from Integral Domains

In analogy to how we proceeded in Theorem 6.5 of Chapter II, in which we extended regular abelian semigroups to abelian groups, we would like in this section to embed integral domains in fields.

Remark 5.1. We recall from Definition 2.9 that an integral domain $(R, +, \cdot)$ is a nonnull commutative ring without zero divisors. This means in particular that nonzero elements $a, b \in R$ satisfy $a \cdot b \neq 0$.

Furthermore, we observe that for an integral domain $(R, +, \cdot)$, we have that $(R \setminus \{0\}, \cdot)$ is a regular abelian semigroup. (Note that since $R \neq \{0\}$, it follows that $R \setminus \{0\}$ is nonempty.) Since there are no zero divisors, it follows that $R \setminus \{0\}$ is closed with respect to multiplication. The commutativity of multiplication is obvious. If for $a, b, c \in R \setminus \{0\}$, we have the equality $a \cdot c = b \cdot c$, we can transform this in $(R, +, \cdot)$ to

$$(a - b) \cdot c = 0.$$

Since $c \neq 0$, it follows that $a - b = 0$, which means that we must have $a = b$. That is, we can "cancel" the c. This proves the regularity of the semigroup $(R \setminus \{0\}, \cdot)$.

Theorem 5.2. *For every integral domain $(R, +, \cdot)$, there exists a uniquely determined field $(K, \oplus, \odot)$ satisfying the following properties:*

(i) *R is a subset of K, and the restrictions of $\oplus$ and $\odot$ to R agree with the operations $+$ and $\cdot$.*

(ii) *If $(K', \oplus', \odot')$ is another field satisfying (i), then K is a subfield of K'.*

Proof. We must prove existence and uniqueness. We begin with the latter.

Uniqueness: The proof of uniqueness of the field $(K, \oplus, \odot)$ to be constructed proceeds analogously to the proof of uniqueness in Theorem 6.5 of Chapter II using property (ii).

Existence: To prove the existence of the field in question, we consider the set

$$M := R \times (R \setminus \{0\}) = \{(a, b) \mid a \in R, b \in R \setminus \{0\}\}$$

with the relation $\sim$ defined as follows:

$$(a, b) \sim (c, d) \quad \Longleftrightarrow \quad a \cdot d = b \cdot c \quad (a, c \in R; b, d \in R \setminus \{0\}).$$

Although the initial situation here is similar to that in the proof of Theorem 6.5 of Chapter II, there is a subtle difference to be considered, namely that the Cartesian product M has an asymmetry due to the fact that the two factors are unequal.

We first observe that as in Theorem 6.5 of Chapter II, the relation $\sim$ is an equivalence relation.

(a) Reflexivity: Since the multiplication is commutative, we have for all $a \in R, b \in R \setminus \{0\}$ the equality $a \cdot b = b \cdot a$. That is, $(a, b) \sim (a, b)$. The relation $\sim$ is therefore reflexive.

(b) Symmetry: Let $(a, b), (c, d) \in M$ have the property $(a, b) \sim (c, d)$. That is, $a \cdot d = b \cdot c$. Since the multiplication is commutative, we conclude that $c \cdot b = d \cdot a$, which means precisely that $(c, d) \sim (a, b)$. That is, $\sim$ is symmetric.

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

$$a \cdot d = b \cdot c, \quad c \cdot f = d \cdot e. \tag{9}$$

If we multiply the left-hand and right-hand sides of the two equalities together, we obtain, taking into account the associativity and commutativity of multiplication, the following equivalent equalities:

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

If $c \neq 0$, then since $d \neq 0$, we have also $d \cdot c \neq 0$, on account of the absence of zero divisors in $(R, +, \cdot)$, and we can cancel $(d \cdot c)$ in the last equality from the right and obtain

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

which implies that $(a, b) \sim (e, f)$. If, on the other hand, we have $c = 0$, then we obtain from (9) that $a = e = 0$, which implies $(a, b) = (0, b) \sim (0, f) = (e, f)$. Therefore, the relation $\sim$ is transitive.

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

$$K := M/\sim .$$

Since the ring $(R, +, \cdot)$ contains at least the zero element 0 and an additional element $h \neq 0$, it follows that the set M contains at least the two distinct equivalence classes $[0, h]$ and $[h, h]$. We now define two operations on the set K of equivalence classes, which we denote by $\oplus$ and $\odot$. If we have $[a, b], [a', b'] \in K$, then we define

$$[a,b] \oplus [a',b'] := [a \cdot b' + a' \cdot b, b \cdot b'],$$
$$[a,b] \odot [a',b'] := [a \cdot a', b \cdot b'].$$

Since these definitions apparently depend on the choice of representatives a, b and a', b' of the equivalence classes $[a,b]$ and $[a',b']$, we must prove that these operations $\oplus$ and $\odot$ are well defined by showing that they are, 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 then must show that

$$[a \cdot b' + a' \cdot b, b \cdot b'] = [c \cdot d' + c' \cdot d, d \cdot d'],$$
$$[a \cdot a', b \cdot b'] = [c \cdot c', d \cdot d'].$$

(d) Proof that $\oplus$ is well defined: Since $(c,d) \in [a,b]$ and $(c',d') \in [a',b']$, we have

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

We therefore compute, using the associativity, commutativity, and distributivity in R,

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

from which follows the asserted equivalence

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

(e) Proof that $\odot$ is well defined: Again, since $(c,d) \in [a,b]$ and $(c',d') \in [a',b']$, we have

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

Multiplying these two equations together yields, with the help of the associativity and commutativity of R,

$$(a \cdot d) \cdot (a' \cdot d') = (b \cdot c) \cdot (b' \cdot c') \quad \Longleftrightarrow \quad (a \cdot a') \cdot (d \cdot d') = (b \cdot b') \cdot (c \cdot c'),$$

from which follows the asserted equivalence

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

In sum, in $(K, \oplus, \odot)$, we have a set containing at least the two distinct elements $[0,h]$, $[h,h]$ and possessing two operations. In the following three steps, we shall show that $(K, \oplus, \odot)$ is a field. We begin with the proof that $(K, \oplus)$ is an abelian group with identity element $[0,h]$.

(1) As we have seen, the set K is nonempty. We leave it to the reader t o prove that the operation $\oplus$ is associative. The commutativity of $\oplus$ can be seen from the calculation, with $[a,b], [a',b'] \in K$,

$$[a, b] \oplus [a', b'] = [a \cdot b' + a' \cdot b, b \cdot b'] = [a' \cdot b + a \cdot b', b' \cdot b] = [a', b'] \oplus [a, b],$$

where we have used the commutativity of $+$ and $\cdot$.
Since $h \neq 0$, we have the equivalent equalities

$$a \cdot b = b \cdot a \quad \Longleftrightarrow \quad (a \cdot b) \cdot h = (b \cdot a) \cdot h \quad \Longleftrightarrow \quad (a \cdot h) \cdot b = (b \cdot h) \cdot a.$$

That is, $(a \cdot h, b \cdot h) \sim (a, b)$. We see, then, that $[0, h]$ is the identity element of $(K, \oplus)$, since we have, for all $[a, b] \in K$,

$$[a, b] \oplus [0, h] = [a \cdot h + 0 \cdot b, b \cdot h] = [a \cdot h, b \cdot h] = [a, b].$$

The additive inverse of the element $[a, b] \in K$ is given by $[-a, b] \in K$, since we have

$$[a, b] \oplus [-a, b] = [a \cdot b - a \cdot b, b \cdot b] = [0, b \cdot b] = [0, h],$$

where we have used the equivalence $(0, b \cdot b) \sim (0, h)$. We have therefore proved that $(K, \oplus)$ is an abelian group with identity element $[0, h]$.

(2) Our second step is to show that $(K \setminus \{[0, h]\}, \odot)$ is an abelian group with identity element $[h, h]$.
As already mentioned, we have $[h, h] \neq [0, h]$. That is, $[h, h] \in K \setminus \{[0, h]\}$, from which we see that $K \setminus \{[0, h]\}$ is nonempty. The associativity of the operation $\odot$ follows at once from the associativity of $\cdot$, namely

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

The proof that $\odot$ is commutative follows just as easily, using the commutativity of $\cdot$. Using again the equality of equivalence classes $[a \cdot h, b \cdot h] = [a, b]$, we further compute

$$[a, b] \odot [h, h] = [a \cdot h, b \cdot h] = [a, b].$$

We see, then, that $[h, h]$ is an identity element for $K \setminus \{[0, h]\}$. Finally, to determine the multiplicative inverse of an element $[a, b] \in K \setminus \{[0, h]\}$, we observe that on account of $(a, b) \nsim (0, h)$, we have also $a \neq 0$, whence we also have $(b, a) \in M$. We now claim that the multiplicative inverse of $[a, b] \in K \setminus \{[0, h]\}$ is given by the element $[b, a]$, which, by what we have just noted, again lies in $K \setminus \{[0, h]\}$. In fact, we have

$$[a, b] \odot [b, a] = [a \cdot b, b \cdot a] = [h, h],$$

since $(a \cdot b) \cdot h = (b \cdot a) \cdot h$. We have thereby proved that $(K \setminus \{[0,h]\}, \odot)$ is an abelian group with identity element $[h,h]$.

(3) To complete the proof of the field properties of $(K, \oplus, \odot)$, we must verify the distributive laws. As an example, we carry out the proof for the validity of one of the two laws. For $[a,b], [a',b'], [a'',b''] \in K$, we calculate

$$\begin{aligned}
[a,b] \odot ([a',b'] \oplus [a'',b'']) &= [a,b] \odot [a' \cdot b'' + a'' \cdot b', b' \cdot b''] \\
&= [a \cdot (a' \cdot b'' + a'' \cdot b'), b \cdot (b' \cdot b'')] \\
&= [a \cdot a' \cdot b'' + a \cdot a'' \cdot b', b \cdot b' \cdot b''] \\
&= [(a \cdot a') \cdot (b \cdot b'') + (a \cdot a'') \cdot (b \cdot b'), (b \cdot b') \cdot (b \cdot b'')] \\
&= [a \cdot a', b \cdot b'] \oplus [a \cdot a'', b \cdot b''] \\
&= [a,b] \odot [a',b'] \oplus [a,b] \odot [a'',b''].
\end{aligned}$$

Altogether, we have proved that $(K, \oplus, \odot)$ is a field with zero element $[0,h]$ and unit element $[h,h]$. To complete the proof, we must show that $(K, \oplus, \odot)$ satisfies the two asserted properties (i), (ii), that is, (i) that R is a subset of K and the restrictions of $\oplus$ and $\odot$ to R coincide with the operations $+$ and $\cdot$, and (ii) that $(K, \oplus, \odot)$ is minimal with respect to property (i).

To prove property (i), it suffices to find an injective mapping $f : R \longrightarrow K$ satisfying

$$f(a+b) = f(a) \oplus f(b) \qquad (a,b \in R), \tag{10}$$

$$f(a \cdot b) = f(a) \odot f(b) \qquad (a,b \in R). \tag{11}$$

If we then identify R with its image $f(R) \subseteq K$, we obtain the desired result using (10) and (11). We define the mapping $f : R \longrightarrow K$ by sending the element $a \in R$ to the element $[a \cdot h, h] \in K$ (the element h was selected in the construction of the unit element $[h,h]$ of K). We now show, to begin with, that f is injective. To this end, let $a, b \in R$ be such that

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

But in consideration of the properties of the integral domain $(R, +, \cdot)$, this is equivalent to

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

from which follows the injectivity of f.

To prove (10), we choose two arbitrary elements $a, b \in R$ and calculate, taking into account the distributivity in $(R, +, \cdot)$,

$$\begin{aligned}
f(a+b) &= [(a+b) \cdot h, h] = [a \cdot h + b \cdot h, h] = [(a \cdot h) \cdot h + (b \cdot h) \cdot h, h \cdot h] \\
&= [a \cdot h, h] \oplus [b \cdot h, h] = f(a) \oplus f(b).
\end{aligned}$$

To prove (11), we choose two arbitrary elements $a, b \in R$ and calculate, taking into account the associativity and commutativity of $\cdot$,

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

We have thus proved the structure-preserving properties of f asserted in (10) and (11), and so we see that $(K, \oplus, \odot)$ is indeed a field that satisfies property (i).

To finish the proof, we show, finally, that the field $(K, \oplus, \odot)$ that we have constructed is minimal. To do so, we begin with the end of the proof of Theorem 6.5 of Chapter II and show that if $[a \cdot h, h] \in K$ for $a \in R$, $a \neq 0$, then we have also $[h, a \cdot h] \in K$, and K as a field must necessarily contain all elements of the form $[a, b]$ for $a \in R$ and $b \in R \setminus \{0\}$, which proves the minimality of K. □

Exercise 5.3. Complete the proof of Theorem 5.2 by proving the associativity of $\oplus$, the commutativity of $\odot$, and the second distributive law.

Definition 5.4. Let $(R, +, \cdot)$ be an integral domain. The field $(K, \oplus, \odot)$ constructed in Theorem 5.2 is called the *field of fractions of* R and is denoted by $\mathrm{Quot}(R)$. The elements $[a, b] \in K$ are usually represented in the form $a \cdot b^{-1}$ or $\frac{a}{b}$ or a/b.

Exercise 5.5. Show that if $(K, +, \cdot)$ is a field, then the construction of the field of fractions produces nothing new; that is, there exists a ring isomorphism $(\mathrm{Quot}(K), \oplus, \odot) \cong (K, +, \cdot)$.

6. The Rational Numbers

We would like now to consider in greater depth the field $(K, \oplus, \odot)$ constructed in Theorem 5.2 from the integral domain $(R, +, \cdot) = (\mathbb{Z}, +, \cdot)$. This will lead us to the field of *rational numbers*.

We note first that the equivalence relation $\sim$ defined on the Cartesian product $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ now takes the form

$$(a, b) \sim (c, d) \quad \Longleftrightarrow \quad a \cdot d = b \cdot c \quad (a, b \in \mathbb{Z}; b, d \in \mathbb{Z} \setminus \{0\}).$$

The field $(K, \oplus, \odot)$ is given, by the proof of Theorem 5.2, by the set of all equivalence classes $\frac{a}{b} = [a, b]$ of pairs $(a, b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ with the operations

$$\frac{a}{b} \oplus \frac{a'}{b'} = \frac{a \cdot b' + a' \cdot b}{b \cdot b'} \quad \text{and} \quad \frac{a}{b} \odot \frac{a'}{b'} = \frac{a \cdot a'}{b \cdot b'},$$

for $\frac{a}{b}, \frac{a'}{b'} \in K$. The zero element of $(K, \oplus, \odot)$ is given by $\frac{0}{1}$ and the unit element by $\frac{1}{1}$, where 0 and 1 denote the respective integers zero and one.

We see from the proof of Theorem 5.2 that the set of integers $\mathbb{Z}$ stands in bijection to the set $\{\frac{a}{1} \mid a \in \mathbb{Z}\}$. This bijection is induced by the assignment $a \mapsto [a \cdot 1, 1] = \frac{a}{1}$. Once we have identified the set of integers $\mathbb{Z}$ with the set $\{\frac{a}{1} \mid a \in \mathbb{Z}\}$, that is, once we set $a = \frac{a}{1}$, we may then consider $\mathbb{Z}$ to be a subset of K.

Definition 6.1. We shall hereinafter denote the set $(K, \oplus, \odot)$ by $(\mathbb{Q}, +, \cdot)$, and we shall call it the *field of rational numbers*. We may represent $\mathbb{Q}$ as a set in the form

$$\mathbb{Q} = \left\{ \frac{a}{b} \,\middle|\, a \in \mathbb{Z},\, b \in \mathbb{Z} \setminus \{0\} \right\}.$$

We call the rational number $\frac{a}{b}$ a *fraction* or the *quotient* of the integers a and b.

Remark 6.2. (i) For fractions $\frac{a}{b}$ and $\frac{a'}{b'}$, we can rediscover the well-known operations of addition, subtraction, and multiplication, namely

$$\frac{a}{b} \pm \frac{a'}{b'} = \frac{a \cdot b' \pm a' \cdot b}{b \cdot b'} \quad \text{and} \quad \frac{a}{b} \cdot \frac{a'}{b'} = \frac{a \cdot a'}{b \cdot b'}.$$

If $\frac{a}{b} \neq 0$, then we have the familiar rule

$$\left(\frac{a}{b}\right)^{-1} = \frac{b}{a}.$$

(ii) The zero element 0 and unit element 1 of the integers $\mathbb{Z}$ are also, by the identification made above, equal to the zero and unit elements of the rational numbers $\mathbb{Q}$.

(iii) In considering the quotient $\frac{a}{b}$, one should always keep in mind the underlying equivalence class, namely

$$\frac{3}{5} = \frac{6}{10} = \frac{9}{15} = \cdots.$$

That is, each of the pairs $(3, 5), (6, 10), (9, 15), \ldots$ is a representative of the rational number $\frac{3}{5}$.

Of course, behind this is lurking the fact that by construction, we have

$$\frac{a}{b} = \frac{c}{d} \quad \Longleftrightarrow \quad a \cdot d = b \cdot c.$$

Exercise 6.3. Show that every rational number r has exactly one representation $(a, b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ such that a and b are relatively prime and $b \in \mathbb{N} \setminus \{0\}$.

Exercise 6.4. Prove that the set $\mathbb{Q}$ of rational numbers is countable; that is, show that there is a bijection between $\mathbb{Q}$ and $\mathbb{N}$ as sets.

Definition 6.5. We extend to the set $\mathbb{Q}$ of rational numbers the relations $<$ and $\leq$ on the set $\mathbb{Z}$ of integers from Definition 7.4 of Chapter II: for two rational numbers $\frac{a}{b}$, $\frac{a'}{b'}$, we set

$$\frac{a}{b} < \frac{a'}{b'} \iff \begin{cases} a \cdot b' < a' \cdot b, & \text{if } b > 0, b' > 0 \text{ or } b < 0, b' < 0, \\ a \cdot b' > a' \cdot b, & \text{if } b > 0, b' < 0 \text{ or } b < 0, b' > 0, \end{cases}$$

and

$$\frac{a}{b} \leq \frac{a'}{b'} \iff \begin{cases} a \cdot b' \leq a' \cdot b, & \text{if } b > 0, b' > 0 \text{ or } b < 0, b' < 0, \\ a \cdot b' \geq a' \cdot b, & \text{if } b > 0, b' < 0 \text{ or } b < 0, b' > 0. \end{cases}$$

We extend the relations $>$ and $\geq$ to $\mathbb{Q}$ analogously.

Remark 6.6. With the relation $<$, the set $\mathbb{Q}$ of rational numbers becomes an *ordered set*; that is, the following three conditions are satisfied:

(i) For elements $\frac{a}{b}, \frac{a'}{b'} \in \mathbb{Q}$, we have $\frac{a}{b} < \frac{a'}{b'}$ or $\frac{a'}{b'} < \frac{a}{b}$ or $\frac{a}{b} = \frac{a'}{b'}$.

(ii) The three relations $\frac{a}{b} < \frac{a'}{b'}$, $\frac{a'}{b'} < \frac{a}{b}$, $\frac{a}{b} = \frac{a'}{b'}$ are mutually exclusive.

(iii) If $\frac{a}{b} < \frac{a'}{b'}$ and $\frac{a'}{b'} < \frac{a''}{b''}$, then $\frac{a}{b} < \frac{a''}{b''}$.

Analogous conditions hold for $>$.

Exercise 6.7. Determine how the rules for addition and multiplication from Remark 1.19 of Chapter I can be extended to the rational numbers and prove their validity.

Definition 6.8. Let $\frac{a}{b} \in \mathbb{Q}$ be a rational number. We set

$$\left|\frac{a}{b}\right| := \begin{cases} a \cdot b^{-1}, & \text{if } a \cdot b^{-1} \geq 0, \\ -a \cdot b^{-1}, & \text{if } a \cdot b^{-1} < 0. \end{cases}$$

We call the rational number $|\frac{a}{b}|$ the *absolute value* of the rational number $\frac{a}{b}$.

7. Unique Factorization Domains, Principal Ideal Domains, and Euclidean Domains

To conclude this chapter, we would like to investigate how we might carry over the theory of divisibility in the ring $(\mathbb{Z}, +, \cdot)$, which we learned about

in Chapter II, to integral domains $(R, +, \cdot)$ with unit 1. We shall place particular emphasis on the notion of greatest common divisor. Throughout this section, we shall let $(R, +, \cdot)$ be an integral domain with unit element 1.

We begin with a generalization of the notion of divisibility from Definition 2.1 of Chapter I.

Definition 7.1. An element $b \in R$, $b \neq 0$, *divides* an element $a \in R$, denoted by $b \mid a$, if there exists an element $c \in R$ such that $a = b \cdot c$. We say also that b is a *divisor* of a. Furthermore, $b \in R$ is a *common divisor* of $a_1, a_2 \in R$ if there exist $c_1, c_2 \in R$ such that $a_j = b \cdot c_j$ for $j = 1, 2$.

We next extend the notions of greatest common divisor and least common multiple from Definitions 4.1 and 4.7 of Chapter I to integral domains with unit element 1.

Definition 7.2. Let a, b be elements of R not both equal to the zero element 0. An element $d \in R$ satisfying the following two properties is called a *greatest common divisor of a and b*:
(i) $d \mid a$ and $d \mid b$, that is, d is a common divisor of a, b;
(ii) for all $x \in R$ with $x \mid a$ and $x \mid b$, we have $x \mid d$, that is, every common divisor of a, b divides d.

Definition 7.3. Let a, b be nonzero elements of R. An element $m \in R$ satisfying the following two properties is called a *least common multiple* of a and b:
(i) $a \mid m$ and $b \mid m$, that is, m is a common multiple of a, b;
(ii) for all $y \in R$ with $a \mid y$ and $b \mid y$, we have $m \mid y$, that is, every common multiple of a, b is a multiple of m.

Exercise 7.4. Determine a greatest common divisor and least common multiple of the polynomials $20X$ and $10X^2 + 4X - 6$ in the polynomial ring $\mathbb{Z}[X]$.

Remark 7.5. Once we leave the familiar territory of the ring of integers, it is no longer clear whether a greatest common divisor of two ring elements even exists. If there is a greatest common divisor d, we know that all of the associates of d, that is, all products $e \cdot d$ with $e \in R^\times$, also satisfy the properties of a greatest common divisor. That is, there is, in general, no sense in speaking about *the* greatest common divisor. An analogous comment holds for the notion of least common multiple.

Finally, we carry over the notion of prime number to integral domains with unit element 1.

Definition 7.6. An element $p \in R \setminus R^\times$, $p \neq 0$, is said to be *irreducible* if it is divisible only by the units of R and its own associates.

An element $a \in R \setminus R^\times$, $a \neq 0$, that is not irreducible is said to be *reducible*.

An element $p \in R \setminus R^\times$, $p \neq 0$, is said to be *prime* if $p \mid a \cdot b$ for $a, b \in R$ implies $p \mid a$ or $p \mid b$.

Remark 7.7. We note without proof that prime elements are always irreducible. We remark, however, that the converse of this statement is in general false.

Example 7.8. In the integral domain $\mathbb{Z}$, the units are ± 1; the irreducible elements are the integers $\pm p$ with p a prime. By Euclid's lemma (Lemma 1.7), the irreducible elements are also prime.

Exercise 7.9. Give some examples of irreducible elements in the polynomial rings $\mathbb{Z}[X]$ and $\mathbb{Q}[X]$. Are these irreducible elements also prime?

Remark 7.10. Recalling the discussion in Chapter I, we may see that the existence of a greatest common divisor is an immediate consequence of the fundamental theorem of arithmetic. As for the question of the existence of greatest common divisors in integral domains, one is led to the question of existence and uniqueness of factorization (up to order of the factors and multiplication by units) of reducible elements in such rings into irreducible factors. As a negative result in this direction, we remark here that later on, we shall see examples of integral domains in which there is no suitable analogue of the fundamental theorem of arithmetic. With this in mind, we shall see that it is useful to extend the notion of divisibility to ideals.

Definition 7.11. Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals of R. The ideal $\mathfrak{b}$ *divides* the ideal $\mathfrak{a}$, denoted by $\mathfrak{b} \mid \mathfrak{a}$, if the two ideals satisfy the inclusion $\mathfrak{b} \supseteq \mathfrak{a}$.

The relationship between divisibility of elements and divisibility of ideals is clarified in the following lemma.

Lemma 7.12. *Let $\mathfrak{a} = (a)$ and $\mathfrak{b} = (b)$ be principal ideals of R. Then one has the equivalence*

$$b \mid a \quad \Longleftrightarrow \quad \mathfrak{b} \mid \mathfrak{a}.$$

Proof. (i) If $b \mid a$, then there exists $c \in R$ such that $a = b \cdot c$, from which follows

$$\mathfrak{a} = (a) = a \cdot R = (b \cdot c) \cdot R \subseteq b \cdot R = (b) = \mathfrak{b}.$$

This shows that $\mathfrak{b} \supseteq \mathfrak{a}$; that is, $\mathfrak{b} \mid \mathfrak{a}$.

(ii) Suppose now that $\mathfrak{b} \mid \mathfrak{a}$, that is, by the above definition, that

$$(b) = \mathfrak{b} \supseteq \mathfrak{a} = (a).$$

Since now we have $a \in (a)$, we have also $a \in (b)$, and so there must exist $c \in R$ with $a = b \cdot c$. This shows that $b \mid a$. □

Definition 7.13. Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals of R. Then we call the set

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

the *sum* of the ideals $\mathfrak{a}$ and $\mathfrak{b}$.

Lemma 7.14. *Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals of R. Then we have the following:*

(i) *The sum $\mathfrak{a} + \mathfrak{b}$ of the ideals $\mathfrak{a}$ and $\mathfrak{b}$ is an ideal of R. It is the smallest ideal of R containing $\mathfrak{a}$ and $\mathfrak{b}$.*

(ii) *The intersection $\mathfrak{a} \cap \mathfrak{b}$ of the ideals $\mathfrak{a}$ and $\mathfrak{b}$ is an ideal of R. It is the largest ideal of R that is contained in the ideals $\mathfrak{a}$ and $\mathfrak{b}$.*

Proof. (i) Using the subgroup criterion, Lemma 2.25 of Chapter II, it is easy to verify that $(\mathfrak{a} + \mathfrak{b}, +)$ is a subgroup of $(R, +)$. For $r \in R$ and $a + b \in \mathfrak{a} + \mathfrak{b}$, we obtain

$$r(a + b) = r \cdot a + r \cdot b \in \mathfrak{a} + \mathfrak{b}.$$

This proves that $\mathfrak{a} + \mathfrak{b}$ is an ideal of R. Since an ideal is a group under addition, it must contain all sums of its elements, and so an ideal containing both $\mathfrak{a}$ and $\mathfrak{b}$ must contain all sums of the form $a + b$ (with $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$). This shows that $\mathfrak{a} + \mathfrak{b}$ is the smallest ideal containing both $\mathfrak{a}$ and $\mathfrak{b}$.

(ii) We leave this part of the proof as an exercise for the reader. □

Exercise 7.15. Carry out part (ii) of the proof of Lemma 7.14.

Remark 7.16. We point out here that the union of two ideals is not, in general, an ideal. For example, the union of the ideals $\mathfrak{a} = 2\mathbb{Z}$ and $\mathfrak{b} = 3\mathbb{Z}$ is not even closed under addition, since we have $2 + 3 = 5 \notin 2\mathbb{Z} \cup 3\mathbb{Z}$.

Lemma 7.14 together with Definition 7.11 motivates the following definition.

Definition 7.17. Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals of R. Then the ideal $\mathfrak{a} + \mathfrak{b}$, the sum of those ideals, is the *greatest common divisor* of the ideals $\mathfrak{a}$ and $\mathfrak{b}$, for which we write $(\mathfrak{a}, \mathfrak{b})$.

The intersection ideal $\mathfrak{a} \cap \mathfrak{b}$ is called the *least common multiple* of the ideals $\mathfrak{a}$ and $\mathfrak{b}$, for which we write $[\mathfrak{a}, \mathfrak{b}]$.

We now consider three types of integral domain $(R, +, \cdot)$ with unit element 1 for which the greatest common divisor of two elements exists. In each case, we will show how to calculate the greatest common divisor. We shall denote the greatest common divisor of $a, b \in R$ by (a, b), as usual. We must keep in mind, however, that (a, b) is defined only up to multiplication by a unit in R. In contrast, the principal ideal $\big((a, b)\big)$ generated by (a, b) in R is uniquely determined.

7.1 Unique Factorization Domains

Definition 7.18. An integral domain $(R, +, \cdot)$ with unit element 1 is called a *unique factorization domain* (that is, a ring with unique prime decomposition) if every nonzero, nonunit element (that is, every a in $R \setminus R^\times$, $a \neq 0$) can be represented uniquely (up to order and multiplication by a unit) as the product of powers of irreducible elements. Unique factorization domains are also sometimes called *factorial rings*.

Example 7.19. (i) The ring $(\mathbb{Z}, +, \cdot)$ is a unique factorization domain by Theorem 1.8.

(ii) The set $\mathbb{Q}[X]$ of polynomials in the variable X with rational coefficients and with the usual addition and multiplication of polynomials, that is, $(\mathbb{Q}[X], +, \cdot)$, is an integral domain with unit element 1. It can be shown that $(\mathbb{Q}[X], +, \cdot)$ is a unique factorization domain.

Lemma 7.20. *Let $(R, +, \cdot)$ be a unique factorization domain. Furthermore, let a, b be elements of R not both equal to the zero element 0 with the unique (up to order and multiplication by units) prime-power decompositions*

$$a = \prod_{\substack{p \in R \\ p \text{ irreducible}}} p^{a_p}, \quad b = \prod_{\substack{p \in R \\ p \text{ irreducible}}} p^{b_p}$$

into irreducible elements. Then a greatest common divisor (a, b) of a and b is

$$(a, b) = \prod_{\substack{p \in R \\ p \text{ irreducible}}} p^{d_p},$$

where $d_p := \min(a_p, b_p)$.

Proof. The proof is completely analogous to the proof of Theorem 4.3 of Chapter I. □

Remark 7.21. Later, we shall see examples of rings that are not unique factorization domains. The following theorem, which we give without proof, suggests why it is not so easy to find integral domains that are not unique factorization domains.

Theorem 7.22 (Gauss's theorem). *If $(R, +, \cdot)$ is a unique factorization domain, then the polynomial ring $(R[X], +, \cdot)$ is also a unique factorization domain.* □

7.2 Principal Ideal Domains

Definition 7.23. An integral domain $(R, +, \cdot)$ with unit element 1 is called a *principal ideal domain* if every ideal of R is a principal ideal, that is, if every ideal $\mathfrak{a}$ of R is generated by some $a \in R$, that is, $\mathfrak{a} = (a)$.

Example 7.24. (i) The ring $(\mathbb{Z}, +, \cdot)$ is a principal ideal domain by Lemma 3.19.

(ii) It can be shown that the polynomial ring $(\mathbb{Q}[X], +, \cdot)$ is also a principal ideal domain. See Example 7.34 below.

Exercise 7.25. We have already seen that $(\mathbb{Z}[X], +, \cdot)$ is not a principal ideal domain. Try to find other such examples.

The following theorem shows the relationship between unique factorization domains and principal ideal domains. We present it without proof.

Theorem 7.26. *Every principal ideal domain is a unique factorization domain.* □

Remark 7.27. The converse is false: $(\mathbb{Z}[X], +, \cdot)$ is a unique factorization domain by Gauss's theorem, but it is not a principal ideal domain.

On the other hand, it can be shown that in principal ideal domains, every irreducible element is also a prime element. Thus in a principal ideal domain, the notions of primality and irreducibility are equivalent.

Lemma 7.28. *Let $(R, +, \cdot)$ be a principal ideal domain. Furthermore, let a, b be elements of R, not both equal to the zero element 0. Then the ideal $(a) + (b)$ is a principal ideal. That is, there exists $d \in R$ such that*

$$(a) + (b) = (d).$$

Then d is a greatest common divisor of a and b. That is, $d = (a, b)$.

Proof. We must show that d satisfies the following two properties:
(i) $d \mid a$ and $d \mid b$,
(ii) for all $x \in R$ with $x \mid a$ and $x \mid b$, we have $x \mid d$.

Proof of (i): Since by construction, $(a) \subseteq (a) + (b) = (d)$ and $(b) \subseteq (a) + (b) = (d)$, we obtain at once using Lemma 7.12 that $d \mid a$ and $d \mid b$.

Proof of (ii): Let $x \in R$ be a common divisor of a and b. Then Lemma 7.12 implies that

$$(a) \subseteq (x) \quad \text{and} \quad (b) \subseteq (x).$$

But then the ideal $(d) = (a) + (b)$ is contained in the principal ideal (x); that is, $(d) \subseteq (x)$. Another application of Lemma 7.12 shows that $x \mid d$. □

Lemma 7.29. *Let $(R, +, \cdot)$ be a principal ideal domain. Furthermore, let a, b be elements of R, not both equal to the zero element 0. Then there exist $x, y \in R$ such that a greatest common divisor (a, b) of a, b is given by*

$$(a, b) = x \cdot a + y \cdot b.$$

Proof. By Lemma 7.28, a greatest common divisor $d = (a, b)$ of a, b is determined by the equation of ideals

$$(d) = (a) + (b).$$

That is, we have in particular that $d \in (a) + (b)$. Since now the elements of the ideal $(a) + (b)$ are given by

$$(a) + (b) = \{a' + b' \mid a' \in (a),\, b' \in (b)\} = \{r \cdot a + s \cdot b \mid r, s \in R\},$$

we have that d is of the form

$$d = x \cdot a + y \cdot b,$$

with $x, y \in R$. □

7.3 Euclidean Domains

Definition 7.30. An integral domain $(R, +, \cdot)$ with unit element 1 is called a *Euclidean domain* if there exists a valuation function $w : R \setminus \{0\} \longrightarrow \mathbb{Q}$ satisfying the following two properties:

(i) *(Division with remainder).* If $a, b \in R$, $b \neq 0$, then there exist $q, r \in R$ such that $a = q \cdot b + r$ with $w(r) < w(b)$ or $r = 0$.

(ii) For every $s \in \mathbb{Q}$, the set

$$W(s) := \{w(a) \mid a \in R \setminus \{0\},\, w(a) < s\}$$

is finite.

Example 7.31. The ring of integers $(\mathbb{Z}, +, \cdot)$ is a Euclidean domain with the valuation function $w : \mathbb{Z} \setminus \{0\} \longrightarrow \mathbb{Q}$ given by the absolute value $w(a) := |a|$ $(a \in \mathbb{Z} \setminus \{0\})$. The validity of property (i) in Definition 7.30 is an immediate consequence of Theorem 1.4, on division with remainder of integers. Property (ii) is satisfied because for a given rational number s, at most finitely many integers have absolute value less than s.

Theorem 7.32. *Every Euclidean domain $(R, +, \cdot)$ is a principal ideal domain.*

Proof. Let $(R, +, \cdot)$ be a Euclidean domain with valuation function $w : R \setminus \{0\} \longrightarrow \mathbb{Q}$. We must show that every ideal $\mathfrak{a} \subseteq R$ is principal. If $\mathfrak{a}$ is the zero

ideal, then $\mathfrak{a} = (0)$, and we are done. We may therefore assume that $\mathfrak{a} \neq (0)$. Therefore, $\mathfrak{a}$ has at least one element $a_0 \neq 0$. Let $w_0 := w(a_0) \in \mathbb{Q}$ be the value of a_0 under w. From property (ii) of Definition 7.30, we have that the set

$$\{w(a) \mid a \in \mathfrak{a} \setminus \{0\},\ w(a) < w_0\}$$

is finite. There exists, therefore, $a \in \mathfrak{a}$, $a \neq 0$, with minimal value $w(a)$. Let now $b \in \mathfrak{a}$ be arbitrary. Using property (i), we divide b by a with remainder, that is, we determine $q, r \in R$ such that

$$b = q \cdot a + r$$

with $r = 0$ or $w(r) < w(a)$. If we had $r \neq 0$, then $r = b - q \cdot a$ would be a nonzero element of $\mathfrak{a}$ with a value $w(r)$ that was strictly less than the value $w(a)$ of a. This contradicts the choice of a, and so we must have $r = 0$. We therefore have $b = q \cdot a$, whence $\mathfrak{a} = (a)$. □

Remark 7.33. We note that in Definition 7.30, we did not need to require the existence of a unit element 1, since the existence is a consequence of the other requirements. Namely, since $(R, +, \cdot)$ is an integral domain, the ideal $\mathfrak{a} = R$ is nontrivial. That is, there exists $a \in R$, $a \neq 0$, with minimal value $w(a) \in \mathbb{Q}$. If we now divide a with remainder by itself, we obtain, as in the proof above, $a = e \cdot a$ for some $e \in R$. Canceling the a yields the desired unit element $e = 1$.

Example 7.34. We show that $(\mathbb{Q}[X], +, \cdot)$ is a Euclidean domain. First of all, $(\mathbb{Q}[X], +, \cdot)$ is an integral domain. If $P \in \mathbb{Q}[X]$ is a nonzero polynomial, we may map P to its degree $\deg(P)$, given as the largest natural number appearing as an exponent in P. We thereby obtain the mapping

$$\deg : \mathbb{Q}[X] \setminus \{0\} \longrightarrow \mathbb{N} \subseteq \mathbb{Q}.$$

Division with remainder of polynomials shows that property (i) in Definition 7.30 is satisfied. The validity of property (ii) can be seen from the fact that there are only finitely many possibilities for the degree of a polynomial to be less than a given rational number.

In the previous subsection, we saw in Lemma 7.29 that in principal ideal domains $(R, +, \cdot)$, one can represent a greatest common divisor d of two elements $a, b \in R$ as a *linear combination* $d = x \cdot a + y \cdot b$. However, beyond the existence of $x, y \in R$, we could say nothing about how to determine those elements. The following theorem clarifies the issue.

Theorem 7.35 (Euclidean algorithm). *Let $(R, +, \cdot)$ be a Euclidean domain. We consider, for $a, b \in R$ with $b \neq 0$, an extended division with remainder that leads to the following:*

$$\begin{aligned}
a &= q_1 \cdot b + r_1, \quad 0 < w(r_1) < w(b) \text{ or } r_1 = 0;\\
b &= q_2 \cdot r_1 + r_2, \quad 0 < w(r_2) < w(r_1) \text{ or } r_2 = 0;\\
r_1 &= q_3 \cdot r_2 + r_3, \quad 0 < w(r_3) < w(r_2) \text{ or } r_3 = 0;\\
&\dots\\
r_{n-2} &= q_n \cdot r_{n-1} + r_n, \quad 0 < w(r_n) < w(r_{n-1}) \text{ or } r_n = 0;\\
r_{n-1} &= q_{n+1} \cdot r_n + r_{n+1}, \quad 0 < w(r_{n+1}) < w(r_n) \text{ or } r_{n+1} = 0;\\
&\dots
\end{aligned}$$

This process ends after finitely many steps; that is, there exists $n \in \mathbb{N}$ such that $r_{n+1} = 0$. Moreover, the last nonvanishing remainder r_n is a greatest common divisor of a and b.

Proof. Since $(R, +, \cdot)$ is a Euclidean domain, the set of values

$$W\big(w(b)\big) := \big\{ w(a) \,\big|\, a \in R \setminus \{0\},\, w(a) < w(b) \big\}$$

is finite. This has the consequence that the extended division with remainder must end after finitely many steps, that is, that there exists $n \in \mathbb{N}$ such that $r_{n+1} = 0$. In what follows, r_n will denote the last nonvanishing remainder.

We now show that $r_n = (a, b)$. For r_n, we have to verify the two properties of a greatest common divisor from Definition 7.2.

(i) We show first that r_n is a common divisor of a and b. If we look at the last line in the display above, we see that $r_n \mid r_{n-1}$. From the penultimate line, $r_{n-2} = q_n \cdot r_{n-1} + r_n$, we conclude that $r_n \mid r_{n-2}$. Working our way upward through successive rows, we obtain $r_n \mid b$ and finally $r_n \mid a$, and so r_n is indeed a common divisor of a and b.

(ii) We show now that r_n divides every common divisor x of a and b. From the first row of the above display, we obtain $x \mid r_1$. We conclude from the second row that $x \mid r_2$. Continuing in this way, we see that x must divide all the successive remainders, that is, we obtain in particular that $x \mid r_n$, as asserted. □

Remark 7.36 *(Extended Euclidean algorithm).* Let $(R, +, \cdot)$ be a Euclidean domain, and $a, b \in R$ with $b \neq 0$. An analysis of Theorem 7.35 shows that from the data of the extended division with remainder, we can determine *explicit* elements $x, y \in R$ such that

$$(a, b) = x \cdot a + y \cdot b.$$

From the penultimate row of the display given in Theorem 7.35, we can read off that $r_n = r_{n-2} - q_n \cdot r_{n-1}$. Using the third row from the end, that is, $r_{n-3} = q_{n-1} \cdot r_{n-2} + r_{n-1}$, we obtain

$$\begin{aligned} r_n &= r_{n-2} - q_n \cdot r_{n-1} \\ &= r_{n-2} - q_n \cdot (r_{n-3} - q_{n-1} \cdot r_{n-2}) \\ &= (-q_n) \cdot r_{n-3} + (1 + q_n \cdot q_{n-1}) \cdot r_{n-2}. \end{aligned}$$

We see, then, that by working our way up the display, we can represent r_n as a linear combination of two successive remainders r_j, r_{j+1} $(j = n-2, \ldots, 1)$. After $n-2$ steps, we obtain

$$r_n = x_1 \cdot r_1 + x_2 \cdot r_2$$

with suitable $x_1, x_2 \in R$. If in this equation we replace r_2 by $b - q_2 \cdot r_1$ and then r_1 by $a - q_1 \cdot b$, we obtain the desired $x, y \in R$.

To close the main part of this chapter, we would like to illustrate some of the concepts presented by means of an example.

Example 7.37. Consider the Euclidean domain $(\mathbb{Z}, +, \cdot)$ of integers. We wish to compute the greatest common divisor (a, b) for $a = 113$ and $b = 29$ and represent it as a linear combination of integers. Repeated division with remainder yields

$$\begin{aligned} 113 &= 3 \cdot 29 + 26, \\ 29 &= 1 \cdot 26 + 3, \\ 26 &= 8 \cdot 3 + 2, \\ 3 &= 1 \cdot 2 + 1, \\ 2 &= 2 \cdot 1 + 0, \end{aligned}$$

whence we have $(113, 29) = 1$. To obtain the desired integer linear combination representing the greatest common divisor, we proceed through the display above from bottom to top. We obtain

$$\begin{aligned} 1 &= 3 - 1 \cdot 2, \\ &= 3 - 1 \cdot (26 - 8 \cdot 3) = 9 \cdot 3 - 1 \cdot 26, \\ &= 9 \cdot (29 - 1 \cdot 26) - 1 \cdot 26 = 9 \cdot 29 - 10 \cdot 26, \\ &= 9 \cdot 29 - 10 \cdot (113 - 3 \cdot 29) = -10 \cdot 113 + 39 \cdot 29, \end{aligned}$$

and we have

$$(113, 29) = 1 = -10 \cdot 113 + 39 \cdot 29.$$

Exercise 7.38. Carry out the Euclidean algorithm to determine the greatest common divisor (a, b) of a, b in the following two cases:

(a) $a = 123\,456\,789$, $b = 555\,555\,555$ in the ring $(\mathbb{Z}, +, \cdot)$.

(b) $a = X^4 + 2X^3 + 2X^2 + 2X + 1$, $b = X^3 + X^2 - X - 1$ in the polynomial ring $(\mathbb{Q}[X], +, \cdot)$.

C. Rational Solutions of Equations: A First Glimpse

We conclude this chapter with a first look at the solution of polynomial equations in rational numbers, which were introduced in Section 6 of this chapter. This will lead us to some classical questions, which in part were resolved only relatively recently and some of which are topics of active research in number theory today.

C.1 The General Problem

In generalizing linear algebra, where one looks for common solutions to several linear equations in several variables $X_1, \ldots, X_n$ in a field K (such as the field of rational numbers $\mathbb{Q}$, the field of real numbers $\mathbb{R}$, and the field of complex numbers $\mathbb{C}$, which will be constructed in the following chapters), in the field of complex algebraic geometry one is interested in understanding the manifold of all common solutions of equations of the form

$$P_j(X_1, \ldots, X_n) = 0 \quad (j = 1, \ldots, r),$$

where $P_j = P_j(X_1, \ldots, X_n)$ are polynomials of arbitrary degree with coefficients in $\mathbb{C}$; that is, $P_1, \ldots, P_r$ are elements of the polynomial ring $\mathbb{C}[X_1, \ldots, X_n]$.

In arithmetic algebraic geometry, one is interested in the analogous question about the field $\mathbb{Q}$ of rational numbers. In particular, one is interested in the following two fundamental questions:

(A) Is there an n-tuple $(x_1, \ldots, x_n)$ of rational numbers such that

$$P_j(x_1, \ldots, x_n) = 0$$

for $j = 1, \ldots, r$?

(B) If the answer to (A) is affirmative, is the number of such n-tuples finite or infinite?

In what follows, we shall assume that question (A) has been answered in the affirmative, and we shall investigate question (B) for the case of two variables $X = X_1, Y = X_2$ and $r = 1$, that is, the case of a polynomial $P = P_1 \in \mathbb{Q}[X, Y]$.

Let, therefore, $P = P(X, Y)$ be a polynomial in the two variables X, Y with rational coefficients. To answer the two questions posed above, we may assume without loss of generality that the coefficients of P are in fact integers. That is, we may assume that P is an element of the polynomial ring $\mathbb{Z}[X, Y]$. In regard to question (A), we are interested in whether there exist rational numbers x, y such that $P(x, y) = 0$. This question can be formulated geometrically as follows. The equation

$$P(X, Y) = 0$$

defines an *algebraic curve* C in the X,Y-plane. The question of rational solutions x,y of the polynomial equation $P(X,Y)=0$ is thereby reduced to the question of points on the curve C that have rational coordinates. To study question (A), we write

$$C(\mathbb{Q}) := \{(x,y) \in \mathbb{Q}^2 \mid P(x,y)=0\}$$

and call this the set of rational points on C. For example, for the unit circle with center at the origin, defined by the equation $X^2+Y^2-1=0$, we see that $(3/5,4/5)$ is a rational point.

Beginning with the algebraic curve C defined by the equation $P(X,Y)=0$, we may reformulate questions (A) and (B) as follows:

(A) Is the set $C(\mathbb{Q})$ nonempty?

(B) If the answer to (A) is affirmative, is $C(\mathbb{Q})$ finite or infinite?

In the following, we shall give a rough answer to question (B), and we shall proceed by studying the polynomial P for increasing values of its degree d.

C.2 Rational Points on Lines and Quadrics

Degree $d=1$: Without loss of generality, we may assume that

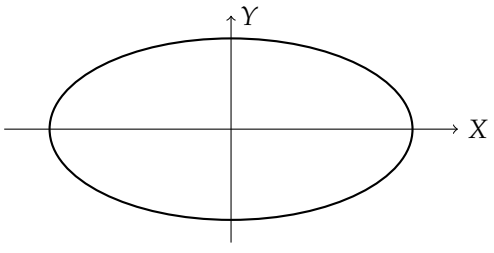
$$P(X,Y) = aX+bY+c$$

with $a,b,c \in \mathbb{Z}$ and $a \neq 0$. Since the curve C defined by $P(X,Y)=0$ is in this case a straight line with rational slope, we see easily that

$$C(\mathbb{Q}) = \left\{ (x,y) \in \mathbb{Q}^2 \,\middle|\, x = -\frac{bt+c}{a}, y = t : t \in \mathbb{Q} \right\},$$

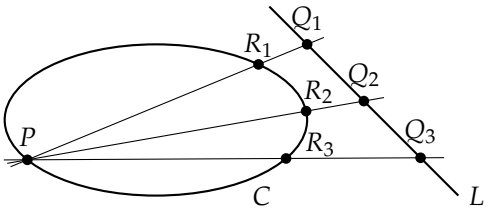
from which we conclude at once that the set $C(\mathbb{Q})$ is always infinite. In particular, it is never empty.

Degree $d=2$: Without loss of generality, we may assume that $P(X,Y) = aX^2+bXY+cY^2+d$ with $a,b,c,d \in \mathbb{Z}$ and $a \neq 0$. The curve C defined by $P(X,Y)=0$ is a conic section, also called a *quadric*.



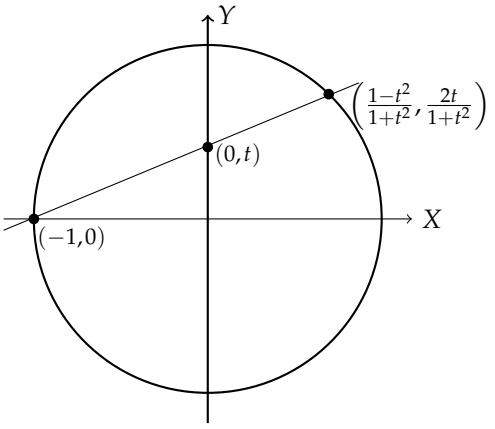
As the example $P(X,Y) = X^2 - 2$ shows, a quadric defined over the rational numbers need not have any rational points. That the curve in this example has no rational points is equivalent to the fact that $\sqrt{2}$ is irrational.

We now assume that the curve C has at least one rational point $P \in C(\mathbb{Q})$. If we draw a line from the point P on C to a rational point Q on a line L with rational slope, that connecting line will also have rational slope, and it will intersect C in an additional point R. The X-coordinate of this point satisfies a quadratic equation one of whose solutions (the X-coordinate of the point P) is rational. Viète's formula tells us that the X-coordinate of the intersection point R, and therefore also its Y-coordinate, must also be rational. As can be seen in the figure below,



we obtain by this method infinitely many rational points on the quadric C, since $L(\mathbb{Q})$ is infinite. We see, then, that the existence of one rational point on a quadric C implies that $C(\mathbb{Q})$ is infinite.

Remark C.1. The qualitative result on the infinitude of rational points on a quadric that we have just proved can be used quantitatively, as shown in the following example. Let the quadric C be the unit circle $X^2 + Y^2 - 1 = 0$, and choose the rational point to be $P = (-1,0)$ and the line L with rational slope to be the Y-axis.



If we join the point $P = (-1,0)$ to the rational point $Q = (0,t)$ on the Y-axis, we obtain a point R on the unit circle with the rational coordinates

$$x = \frac{1-t^2}{1+t^2}, \quad y = \frac{2t}{1+t^2}.$$

Setting $t = n/m$ ($m,n \in \mathbb{N}$, $m > n > 0$), we obtain as a lovely auxiliary result the fact that there are infinitely many triples of natural numbers (a,b,c) such that $a^2 + b^2 = c^2$. They are given by

$$a = m^2 - n^2, \quad b = 2mn, \quad c = m^2 + n^2.$$

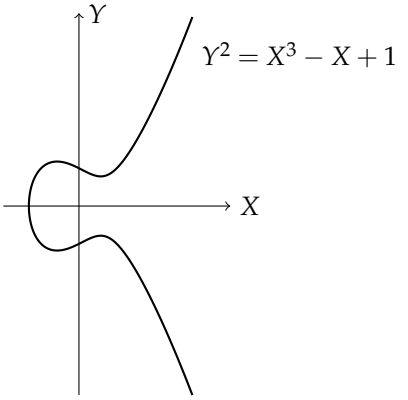
Such triples are called *Pythagorean triples*.

C.3 Rational Points on Elliptic Curves

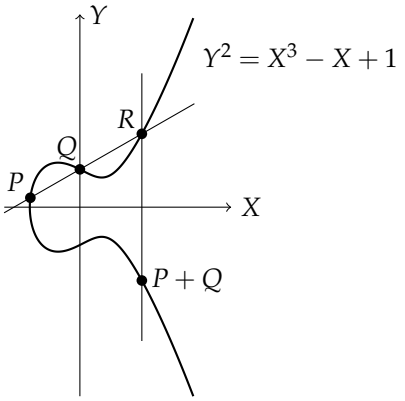
Now let $P = P(X,Y)$ be a polynomial of degree $d = 3$, and C the curve defined by $P(X,Y) = 0$. As in the case of quadrics, the set $C(\mathbb{Q})$ can be empty. We assume in what follows that the curve C has at least one rational point. If we take this point to be the point at infinity on C, we can express the curve C without loss of generality in the form

$$Y^2 = X^3 + aX^2 + bX + c \tag{12}$$

with $a,b,c \in \mathbb{Z}$. If we also assume that the cubic polynomial on the right-hand side of this equation has no multiple roots, that is, that its discriminant Δ does not vanish, then C is called an *elliptic curve*. For more on the theory of elliptic curves, we refer the reader to the textbooks [8] and [14]. We shall here investigate the question whether the set of rational points on an elliptic curve is finite or infinite.



We note first that the set $C(\mathbb{Q})$ of rational points of C has the structure of an abelian group, with the group operation defined as follows. The sum $P+Q$ of two rational points $P,Q \in C(\mathbb{Q})$ is given by the following rational point: Join points P and Q by a straight line L. This line has rational slope and therefore intersects the cubic C in some rational point R. We define the reflection of R in the X-axis, which is a rational point of C, to be the sum $P+Q \in C(\mathbb{Q})$.



This construction shows at once that the operation of addition thus defined is commutative. It is not so easy, however, to show that addition is associative. The point at infinity is the identity element of the abelian group $C(\mathbb{Q})$.

In 1922, the English mathematician Louis Mordell determined the structure of the abelian group $C(\mathbb{Q})$.

Theorem C.2 (Mordell [13]). *If C is an elliptic curve defined over the rational numbers, then the abelian group $C(\mathbb{Q})$ is finitely generated. Thus the group has the direct sum decomposition*

$$C(\mathbb{Q}) = C(\mathbb{Q})_{\text{free}} \oplus C(\mathbb{Q})_{\text{tors}},$$

where $C(\mathbb{Q})_{\text{free}}$ is the free part, and $C(\mathbb{Q})_{\text{tors}}$ the finite part, called the torsion subgroup, of the abelian group $C(\mathbb{Q})$. □

The set $C(\mathbb{Q})_{\text{tors}}$ is a finite abelian group; that is, $C(\mathbb{Q})_{\text{tors}}$ consists of the rational points of C of finite order.

Theorem C.3 (Mazur [12]). *If C is an elliptic curve defined over the rational numbers, then the torsion subgroup $C(\mathbb{Q})_{\text{tors}}$ is isomorphic to one of the following* 15 *groups:*

$$\mathbb{Z}/N\mathbb{Z} \quad (N = 1,\dots,10,12),$$
$$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2N\mathbb{Z} \quad (N = 1,\dots,4).$$

□

For the free part, one has the isomorphism

$$C(\mathbb{Q})_{\text{free}} \cong \mathbb{Z}^{r_C} = \underbrace{\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}_{r_C \text{ times}}.$$

The number r_C is called the *rank* of $C(\mathbb{Q})$. If $r_C = 0$, then $C(\mathbb{Q})$ has only finitely many rational points. If, on the other hand, we have $r_C > 0$, then $C(\mathbb{Q})$ has rational points of infinite order and therefore infinitely many rational points. In sum, we have

$$\begin{aligned} r_C = 0 \quad &\Longleftrightarrow \quad \#C(\mathbb{Q}) < \infty, \\ r_C > 0 \quad &\Longleftrightarrow \quad \#C(\mathbb{Q}) = \infty. \end{aligned}$$

The problem of describing the group $C(\mathbb{Q})$ of elliptic curves C consists, therefore, essentially in determining its rank r_C.

C.4 The Conjecture of Birch and Swinnerton-Dyer

As before, let C be an elliptic curve defined by an equation of the form (12). The conjecture of Birch and Swinnerton-Dyer provides an analytic tool to decide whether $r_C = 0$ or $r_C > 0$. To formulate the conjecture, we now consider (12) as a congruence modulo an arbitrary prime number $p \in \mathbb{P}$ and define the quantity

$$N_p := \#\big\{x, y \in \{0, \ldots, p-1\} \mid y^2 \equiv x^3 + ax^2 + bx + c \mod p\big\} + 1.$$

In [2], Bryan Birch and Peter Swinnerton-Dyer gave experimental evidence for the equivalence

$$r_C > 0 \quad \Longleftrightarrow \quad \prod_{\substack{p \in \mathbb{P} \\ p \leq x}} \frac{N_p}{p} \xrightarrow[x \to \infty]{} \infty. \tag{13}$$

Using the L-series $L_C(s)$ of the elliptic curve C, which for $s \in \mathbb{C}$ with $\operatorname{Re}(s) > 3/2$ is defined by the convergent Euler product

$$L_C(s) := \prod_{\substack{p \in \mathbb{P} \\ p \nmid 2\Delta}} \frac{1}{1 - (p + 1 - N_p)p^{-s} + p^{1-2s}},$$

one can rewrite (13), at least formally, as the equivalence

$$r_C > 0 \quad \Longleftrightarrow \quad L_C(1) = 0.$$

After these preliminaries, we are in a position to formulate the conjecture of Birch and Swinnerton-Dyer.

Conjecture (Birch–Swinnerton-Dyer conjecture [2]). *Let C be an elliptic curve defined by* (12). *Then we have the following:*

(i) *The L-series $L_C(s)$ of C can be analytically continued to an entire function of the complex plane $\mathbb{C}$. In particular $L_C(s)$ is defined at the point $s = 1$.*

(ii) *For the order $\operatorname{ord}_{s=1} L_C(s)$ of vanishing of $L_C(s)$ at the point $s = 1$, one has the equality $r_C = \operatorname{ord}_{s=1} L_C(s)$. Moreover, there is an explicit formula that relates the first nonvanishing coefficient of the Taylor development of $L_C(s)$ at $s = 1$ to the arithmetic of the curve C.*

Aside from a few special cases, this conjecture has been proved essentially only for elliptic curves of ranks 0 and 1. More precisely, the following results are known. In 1977, John Coates and Andrew Wiles proved in [4] the finiteness of $C(\mathbb{Q})$ for elliptic curves $C/\mathbb{Q}$ with complex multiplication and $L_C(1) \neq 0$. In 1986, Benedict Gross and Don Zagier proved in [7] that (modular) elliptic curves $C/\mathbb{Q}$ with $L_C(1) = 0$ but $L'_C(1) \neq 0$ have infinitely many rational points. Using this result and some new ideas, Victor Alexandrovich Kolyvagin proved in 1989, in [9], that $L_C(1) \neq 0$ implies $r_C = 0$, and that $L_C(1) = 0$, $L'_C(1) \neq 0$ implies $r_C = 1$. His proof required an analytic assumption that was shortly afterward proved by Daniel Bump, Solomon Friedberg, and Jeffrey Hoffstein in [3]. We refer the reader to the survey article [17] by Andrew Wiles.

The most current results on the Birch–Swinnerton-Dyer conjecture can be found in the fundamental works of Manjul Bhargava, for which he was awarded the Fields Medal in 2014, the highest honor that can be bestowed on a mathematician. Together with Christopher Skinner and Wei Zhang, he proved in [1] that more than 66% of elliptic curves defined over $\mathbb{Q}$ satisfy the Birch–Swinnerton-Dyer conjecture.

The search for rational points on elliptic curves is related to the classical *congruent number problem*, which can be formulated simply as follows: Let F be a positive natural number. Is there a right triangle with rational side lengths a, b, c and area F? That is, do there exist positive rational numbers a, b, c that satisfy the equations

$$a^2 + b^2 = c^2, \quad \frac{a \cdot b}{2} = F\,?$$

The Pythagorean triple $(3,4,5)$ shows, for example, that for $F = 6$, there exists such a triangle, in fact, one with not merely rational, but integral, side lengths. If there exists such a rational right triangle for a given positive natural number F, then we call F a *congruent number*. The *congruent number problem* is to determine whether a given number F is a congruent number.

The congruent number problem is related to the search for rational points on elliptic curves as follows. We associate the positive natural number F

with the elliptic curve

$$C_F : Y^2 = X^3 - F^2X = X(X-F)(X+F).$$

It can be shown that F is a congruent number if and only if the rank r_{C_F} of C_F is positive, which by the Birch–Swinnerton-Dyer conjecture is equivalent to the vanishing of $L_{C_F}(1)$. If $r_{C_F} > 0$, then there exists a rational point $(x,y) \in C_F(\mathbb{Q})$ with $y \neq 0$ (since $(x,0)$ would be a 2-torsion point on C_F). We note, however, that conversely, if $(x,y) \in C_F(\mathbb{Q})$ is a rational point with $y \neq 0$, then (x,y) has infinite order, since all rational points of finite order on C_F can be shown to be 2-torsion points and hence satisfy $y = 0$. We may therefore assume without loss of generality that $x < 0$ and $y > 0$, and we obtain the side lengths of the desired right triangle with area F in the form

$$a = \frac{F^2 - x^2}{y}, \quad b = -\frac{2xF}{y}, \quad c = \frac{F^2 + x^2}{y}.$$

Example C.4. For $F = 5$, the rational point $(-5/9, 100/27) \in C_F(\mathbb{Q})$ has infinite order. We therefore obtain the right triangle with side lengths $a = 20/3$, $b = 3/2$, $c = 41/6$ and area $F = 5$.

For $F = 6$, we obtain for $(-3,9) \in C_F(\mathbb{Q})$ the familiar right triangle with sides of length $a = 3$, $b = 4$, $c = 5$.

For $F = 1,2,3$, it can be shown that $r_{C_F} = 0$, and so $1,2,3$ are not congruent numbers. Since $F = 1$ is not a congruent number, it follows that there is no right triangle with rational side lengths and area equal to the square of an integer.

Some known results: If F satisfies one of the congruences $F \equiv 5,6,7 \mod 8$, then, as shown in [7], F is a congruent number if $L'_{C_F}(1) \neq 0$. This condition is satisfied, for instance, if F is a prime number with $F \equiv 5,7 \mod 8$ (in the latter case, $2F$ is also a congruent number). The most current result as this book goes to press is that of Ye Tian, who showed in [16] that for every positive natural number k, there exist infinitely many square-free congruent numbers F with exactly k distinct prime factors in each of the residue classes $F \equiv 5,6,7 \mod 8$.

If, on the other hand, we have $F \equiv 1,2,3 \mod 8$, it is conjectured that F is not a congruent number. This conjecture has been proved in the case that F is prime with $F \equiv 3 \mod 8$ and in a number of additional cases, thanks to the latest work of Ye Tian, Xinyi Yuan, and Shouwu Zhang.

C.5 Rational Points on Curves of Degree $d > 3$: Fermat's Conjecture

We devote the last section of our *tour d'horizon* to answering our question (B) for algebraic curves C defined by a polynomial $P \in \mathbb{Z}[X,Y]$ of degree $d > 3$.

To simplify the presentation, we assume that the curve C is nonsingular, that is, that there exist no points $(x,y) \in C$ such that

$$\frac{\partial P}{\partial X}(x,y) = 0 \quad \text{and} \quad \frac{\partial P}{\partial Y}(x,y) = 0$$

hold simultaneously. If we furthermore add the points at infinity to C and assume that the curve is nonsingular at those points as well, we obtain a smooth plane projective curve of degree $d > 3$, which we again denote by C. Louis Mordell conjectured in his work [13] mentioned above that in this case, $C(\mathbb{Q})$ is finite. In 1983, a good sixty years later, Gerd Faltings published a proof of this conjecture that earned him a Fields Medal.

Theorem C.5 (Faltings's theorem [6]). *For a smooth plane projective curve C of degree $d > 3$ defined over the rational numbers, the set of rational points is finite.* □

Remark C.6. Faltings's theorem applies not only to plane algebraic curves. To formulate the theorem in its general form, we recall that associated with every smooth projective curve C is a natural number g_C called the genus of the curve. In the case of a smooth plane projective curve C of degree d, the genus g_C is given by the formula

$$g_C = \frac{(d-1)(d-2)}{2}.$$

The general form of Faltings's theorem states that for a smooth projective curve C of genus $g_C > 1$ defined over the rational numbers $\mathbb{Q}$ or over any algebraic number field K, the set of $\mathbb{Q}$-rational, respectively K-rational, points is finite.

Example C.7. A well-known example is the curve C_d defined by

$$P(X,Y) = X^d + Y^d - 1$$

with $d > 3$. Faltings's theorem states that each curve C_d has only finitely many rational points.

Fermat's famous conjecture from the seventeenth century tightens Faltings's finiteness result for this example, since it states that

$$C_d(\mathbb{Q}) = \begin{cases} \{(1,0),(0,1)\}, & \text{for } d \text{ odd}, \\ \{(\pm 1,0),(0,\pm 1)\}, & \text{for } d \text{ even}. \end{cases}$$

A complete proof of Fermat's conjecture, however, was given only in 1995, by Andrew Wiles.

Theorem C.8 (Wiles's theorem [18]). *For $d > 2$, the equation*

$$X^d + Y^d = Z^d$$

has no integer solutions x, y, z with $xyz \neq 0$. □

For an overview of Wiles's proof of Fermat's last theorem and significant contributions by other mathematicians, see the articles [10] and [11]. A more advanced treatment can be found in [5].

This completes our brief look at answers to question (B) in the search for rational solutions to polynomial equations in *two* variables, which has provided a glimpse at the very rich arithmetic results that such a search yields. We also have seen that many questions on this topic remain unanswered.

The search for answers to questions (A) and (B) in the general case of arbitrary systems of polynomials in several variables is a topic of current research. In this case as well, one hopes that the answers will provide interesting arithmetic insights.

References

[1] M. Bhargava, C. Skinner, W. Zhang: *A majority of elliptic curves over $\mathbb{Q}$ satisfy the Birch and Swinnerton-Dyer conjecture.* Preprint, July 17, 2014. Available online at `arXiv:1407.1826`.

[2] B. Birch, H. P. F. Swinnerton-Dyer: *Notes on elliptic curves I, II.* J. Reine Angew. Math. **212** (1963), 7–25; **218** (1965), 79–108.

[3] D. Bump, S. Friedberg, J. Hoffstein: *Non-vanishing theorems for L-functions of modular forms and their derivatives.* Invent. Math. **102** (1990), 543–618.

[4] J. Coates, A. Wiles: *On the conjecture of Birch and Swinnerton-Dyer.* Invent. Math. **39** (1977), 223–251.

[5] G. Cornell, J. H. Silverman, G. Stevens (eds.): *Modular forms and Fermat's last theorem.* Springer, Berlin Heidelberg New York, 1997.

[6] G. Faltings: *Endlichkeitssätze für abelsche Varietäten.* Invent. Math. **73** (1983), 349–366.

[7] B. Gross, D. Zagier: *Heegner points and derivatives of L-series.* Invent. Math. **84** (1986), 225–320.

[8] A.W. Knapp: *Elliptic curves.* Math. Notes 40, Princeton University Press, Princeton, New Jersey, 1992.

[9] V. A. Kolyvagin: *On the Mordell–Weil and Shafarevich–Tate groups for elliptic Weil curves.* Math. USSR, Izv. **33** (1989), 473–499.

[10] J. Kramer: *Über den Beweis der Fermat-Vermutung I, II.* Elem. Math. **50** (1995), 12–25; **53** (1998), 45–60.

[11] J. Kramer: *Fermat's last theorem – the solution of a 300 year old problem.* In: M. Aigner, E. Behrends (eds.), Mathematics Everywhere, 175–183. American Mathematical Society, Providence, Rhode Island, 2010.

[12] B. Mazur: *Modular curves and the Eisenstein ideal.* Publ. Math. IHES **47** (1977), 33–186.

[13] L. J. Mordell: *On the rational solutions of the indeterminate equations of the third and fourth degrees.* Proc. Cambridge Philos. Soc. **21** (1922), 179–192.

[14] J. H. Silverman, J. Tate: *Rational points on elliptic curves.* Springer, Cham, 2nd edition, 2015.

[15] I. Stewart, D. Tall: *Algebraic number theory and Fermat's last theorem.* Chapman and Hall/CRC, 4th edition, 2015.

[16] Y. Tian: *Congruent numbers and Heegner points.* Cambridge J. Math. **2** (2014), 117–161.

[17] A. Wiles: *The Birch and Swinnerton-Dyer conjecture.* Available online at `www.claymath.org/sites/default/files/birchswin.pdf`.

[18] A. Wiles: *Modular elliptic curves and Fermat's last theorem.* Ann. Math. **141** (1995), 443–551.

IV The Real Numbers

1. Decimal Representation of Rational Numbers

Let a be a nonzero natural number. At the end of Chapter I, we used repeated division with remainder to represent a in the form

$$a = \sum_{j=0}^{\ell} q_j \cdot 10^j, \tag{1}$$

with natural numbers $0 \leq q_j \leq 9$ $(j = 0, \ldots, \ell)$ and $q_\ell \neq 0$. For the sum (1), we introduced the decimal representation

$$a = q_\ell q_{\ell-1} \cdots q_1 q_0.$$

This decimal notation can be carried over easily to the integers. Namely, if the integer a is negative, then $a = -|a|$. Thus the decimal representation of the natural number $|a|$ gives us the decimal representation of a in the form

$$a = -q_\ell q_{\ell-1} \cdots q_1 q_0,$$

again with natural numbers $0 \leq q_j \leq 9$ $(j = 0, \ldots, \ell)$ and $q_\ell \neq 0$.

We wish now to extend decimal representation to the rational numbers. To this end, let $\frac{a}{b}$ represent a rational number; that is, $a, b \in \mathbb{Z}$ and $b \neq 0$. Without loss of generality, we may assume that $b > 0$. Using division with remainder for integers, we find for a, b integers q, r with $0 \leq r < b$ such that

$$a = q \cdot b + r \quad \Longleftrightarrow \quad \frac{a}{b} = q + \frac{r}{b}.$$

For the integers q, we have the decimal representation

$$q = \pm \sum_{j=0}^{\ell} q_j \cdot 10^j = \pm q_\ell q_{\ell-1} \ldots q_1 q_0.$$

We now consider the decimal representation of the rational number $0 \leq \frac{r}{b} < 1$. We assume even that $0 < \frac{r}{b} < 1$. We rewrite this as

$$\frac{r}{b} = \frac{1}{10} \cdot \frac{10 \cdot r}{b} \tag{2}$$

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0_4

and divide $10 \cdot r$ with remainder by b. We thereby obtain natural numbers q_{-1}, r_{-1} with $0 \leq r_{-1} < b$ such that

$$10 \cdot r = q_{-1} \cdot b + r_{-1} \quad \Longleftrightarrow \quad \frac{10 \cdot r}{b} = q_{-1} + \frac{r_{-1}}{b}. \tag{3}$$

From the inequality $\frac{r}{b} < 1$, we estimate

$$0 \leq q_{-1} = \frac{10 \cdot r}{b} - \frac{r_{-1}}{b} < \frac{10 \cdot r}{b} < 10,$$

that is, $0 \leq q_{-1} \leq 9$. Substituting (3) into (2) leads to

$$\frac{r}{b} = \frac{1}{10} \cdot \frac{10 \cdot r}{b} = \frac{1}{10}\left(q_{-1} + \frac{r_{-1}}{b}\right) = \frac{q_{-1}}{10} + \frac{1}{10} \cdot \frac{r_{-1}}{b} = \frac{q_{-1}}{10} + \frac{1}{10^2} \cdot \frac{10 \cdot r_{-1}}{b}.$$

If $r_{-1} \neq 0$, we divide $10 \cdot r_{-1}$ with remainder by b and obtain natural numbers q_{-2}, r_{-2} with $0 \leq r_{-2} < b$ such that

$$10 \cdot r_{-1} = q_{-2} \cdot b + r_{-2} \quad \Longleftrightarrow \quad \frac{10 \cdot r_{-1}}{b} = q_{-2} + \frac{r_{-2}}{b}.$$

As before, we estimate $0 \leq q_{-2} \leq 9$ and putting everything together, obtain

$$\begin{aligned} \frac{r}{b} &= \frac{q_{-1}}{10} + \frac{1}{10^2} \cdot \frac{10 \cdot r_{-1}}{b} = \frac{q_{-1}}{10} + \frac{1}{10^2}\left(q_{-2} + \frac{r_{-2}}{b}\right) \\ &= \frac{q_{-1}}{10} + \frac{q_{-2}}{10^2} + \frac{1}{10^3} \cdot \frac{10 \cdot r_{-2}}{b}. \end{aligned}$$

Proceeding, we obtain natural numbers q_{-3}, r_{-3} with $0 \leq q_{-3} \leq 9$ and $0 \leq r_{-3} < b$ such that

$$\frac{r}{b} = \frac{q_{-1}}{10} + \frac{q_{-2}}{10^2} + \frac{q_{-3}}{10^3} + \frac{1}{10^4} \cdot \frac{10 \cdot r_{-3}}{b}.$$

After k steps, we obtain natural numbers q_{-k}, r_{-k} with $0 \leq q_{-k} \leq 9$ and $0 \leq r_{-k} < b$ such that

$$\frac{r}{b} = \sum_{j=1}^{k} \frac{q_{-j}}{10^j} + \frac{1}{10^{k+1}} \cdot \frac{10 \cdot r_{-k}}{b}.$$

With this procedure there are two possibilities: either there exists $k \in \mathbb{N}$, $k > 0$, such that $r_{-k} = 0$, or the remainders r_{-j} are nonzero for all $j = 1, 2, 3, \ldots$.

Definition 1.1. With notation as above, we define the following for $a, b \in \mathbb{Z}$ and $b \neq 0$:

(i) If $r = 0$ or there exists $k \in \mathbb{N}$, $k > 0$, with $r_{-k} = 0$, we set

$$\pm q_\ell \dots q_0 . q_{-1} \dots q_{-k} := \pm \sum_{j=-\ell}^{k} \frac{q_{-j}}{10^j}$$

and call $\pm q_\ell \dots q_0 . q_{-1} \dots q_{-k}$ the *decimal representation* or *decimal expansion* of the rational number $\frac{a}{b}$.

(ii) If all the r_{-j} are nonzero, we set, formally,

$$\pm q_\ell \dots q_0 . q_{-1} \dots q_{-k} \dots := \pm \sum_{j=-\ell}^{\infty} \frac{q_{-j}}{10^j}$$

and call $\pm q_\ell \dots q_0 . q_{-1} \dots q_{-k} \dots$ the *decimal representation* or *decimal expansion* of the rational number $\frac{a}{b}$.

Remark 1.2. We note that the infinite sum (series) in Definition 1.1 (ii),

$$\pm \sum_{j=-\ell}^{\infty} \frac{q_{-j}}{10^j} = \pm \left(q_\ell \cdot 10^\ell + \dots + q_0 + \frac{q_{-1}}{10} + \frac{q_{-2}}{10^2} + \dots \right),$$

makes no sense at the present moment. It is merely a symbolic notation. In contrast, the finite sum in Definition 1.1 (i),

$$\pm \sum_{j=-\ell}^{k} \frac{q_{-j}}{10^j} = \pm \left(q_\ell \cdot 10^\ell + \dots + q_0 + \frac{q_{-1}}{10} + \dots + \frac{q_{-k}}{10^k} \right),$$

has a concrete significance and takes on the value $\frac{a}{b}$; that is, we have by construction that

$$\frac{a}{b} = \pm q_\ell \dots q_0 . q_{-1} \dots q_{-k}.$$

Definition 1.3. We call a nonterminating decimal expansion

$$\pm q_\ell \dots q_0 . q_{-1} \dots q_{-k} \dots$$

periodic if there exist natural numbers $v \geq 0$ and $p > 0$ such that $q_{-(v+j)} = q_{-(v+j+p)} = q_{-(v+j+2p)} = \cdots$ for $j = 1, \dots, p$. For brevity, we write this as follows:

$$\pm q_\ell \dots q_0 . q_{-1} \dots q_{-v} \overline{q_{-(v+1)} \cdots q_{-(v+p)}}.$$

If $v = 0$, the decimal expansion is said to be *purely periodic*. The smallest natural number p as defined above is called the *period* of the decimal expansion of the rational number $\frac{a}{b}$.

Proposition 1.4. *Let $a, b \in \mathbb{Z}$, $b \neq 0$. If the decimal expansion of $\frac{a}{b}$ does not terminate, then it is periodic.*

Proof. Suppose the number $\frac{a}{b}$ has a nonterminating decimal expansion. Looking at the construction of the decimal expansion of $\frac{a}{b}$, we see that the infi-

nite collection of remainders $r_0 := r, r_{-1}, r_{-2}, r_{-3}, \dots$ is actually a subset of $\{0, \dots, b-1\}$. Therefore, there must be at least two remainders r_{-j_1}, r_{-j_2} that are identical. We may assume without loss of generality that $j_2 > j_1 \geq 0$ and for fixed j_1, choose the difference $p := j_2 - j_1$ to be minimal. The algorithm for obtaining the decimal expansion of $\frac{a}{b}$ then shows that

$$\begin{aligned} r_{-j_1} &= r_{-(j_1+p)} = r_{-(j_1+2p)} = \cdots, \\ r_{-(j_1+1)} &= r_{-(j_1+1+p)} = r_{-(j_1+1+2p)} = \cdots, \\ &\cdots \\ r_{-(j_1+p-1)} &= r_{-(j_1+2p-1)} = r_{-(j_1+3p-1)} = \cdots. \end{aligned}$$

If we now choose j_1 to be minimal and set $v := j_1 \geq 0$, we obtain the assertion of the theorem. □

Remark 1.5. The following questions arise:

(i) Is there a set of numbers in which the formal infinite sum

$$\pm \sum_{j=-\ell}^{\infty} \frac{q_{-j}}{10^j}$$

can be given a precise meaning?

(ii) Is there a set of numbers in which *arbitrary* infinite decimal expansions, that is, those that are not necessarily periodic, have a well-defined meaning, that is, in which infinite sums like

$$\pm \sum_{j=-\ell}^{\infty} \frac{q_{-j}}{10^j}$$

represent well-defined numbers?

Exercise 1.6.

(a) Determine the decimal expansions of $\frac{1}{5}$, $\frac{1}{3}$, $\frac{1}{16}$, $\frac{1}{11}$, and $\frac{1}{7}$.

(b) Formulate a criterion for when a fraction $\frac{a}{b}$ ($a, b \in \mathbb{Z}; b \neq 0$) possesses a terminating decimal expansion.

(c) Find a bound on the maximal period length of the decimal expansion of a rational number as a function of its denominator. Give examples for which the period is maximal (with respect to this bound).

(d) Describe a process for obtaining the fraction $\frac{a}{b}$ from its periodic decimal expansion. Use this process on the periodic decimal fraction $0.\overline{123}$.

2. Construction of the Real Numbers

In Chapter I, we constructed the set $\mathbb{N}$ of natural numbers on the basis of the Peano axioms, and we defined on that set the operations of addition and multiplication, which satisfy the commutative, associative, and distributive laws. In particular, we obtained $(\mathbb{N}, +)$ as a regular abelian semigroup, which we then, in Chapter II, extended to the abelian group $(\mathbb{Z}, +)$ of integers. By carrying over the multiplicative structure of the natural numbers to the integers, we obtained, at the beginning of Chapter III, the integral domain $(\mathbb{Z}, +, \cdot)$ of integers. We extended this at the end of Chapter III to the field $(\mathbb{Q}, +, \cdot)$ of rational numbers. In the previous section, we have seen that the decimal expansion of a rational number either terminates or is periodic, and that led us to wonder whether there exists a set of numbers containing numbers with infinite aperiodic decimal expansions. We shall now answer that question in the affirmative. In doing so, we will be led to the construction of the real numbers. We begin with the definition of a *Cauchy sequence*.

But first let us specify a few notational conventions. In what follows, we shall use Latin letters to denote rational numbers, which will be distinguished from the real numbers, which we have yet to introduce, which will be denoted by Greek letters. The one exception will be the Greek letter epsilon, which in the set of rational numbers will be denoted by ϵ, and in the set of real numbers, by ε.

Definition 2.1. A sequence $(a_n) = (a_n)_{n\geq 0}$ with $a_n \in \mathbb{Q}$ for all $n \in \mathbb{N}$ is called a *rational Cauchy sequence* if for every $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, there exists $N(\epsilon) \in \mathbb{N}$ such that for all $m, n \in \mathbb{N}$ with $m, n > N(\epsilon)$, we have the inequality

$$|a_m - a_n| < \epsilon.$$

A sequence $(a_n) = (a_n)_{n\geq 0}$ with $a_n \in \mathbb{Q}$ for all $n \in \mathbb{N}$ is called a *rational null sequence* if for every $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, there exists $N(\epsilon) \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n > N(\epsilon)$, we have the inequality

$$|a_n| < \epsilon.$$

Exercise 2.2.

(a) Prove that the sequences $\left(\frac{1}{n+1}\right)_{n\geq 0}$ and $\left(\frac{n}{2^n}\right)_{n\geq 0}$ are rational null sequences.

(b) Give further examples of rational null sequences.

Remark 2.3. (i) A rational null sequence (a_n) is, in particular, a rational Cauchy sequence. Namely, given $\epsilon/2 \in \mathbb{Q}$, $\epsilon > 0$, there exists $N(\epsilon/2) \in \mathbb{N}$ such that for all $m, n \in \mathbb{N}$ with $m, n > N(\epsilon/2)$, we have the inequality

$$|a_n| < \frac{\epsilon}{2}.$$

Using the triangle inequality, we thereby obtain for $m, n > N(\epsilon/2)$ that

$$|a_m - a_n| \leq |a_m| + |a_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus (a_n) is a rational Cauchy sequence.

(ii) Every rational Cauchy sequence (a_n) is bounded, since for $\epsilon = 1$ and the associated $N(1) \in \mathbb{N}$, which exists by the definition of a Cauchy sequence, we have for all $m, n \in \mathbb{N}$ with $m, n > N(1)$, the inequality

$$|a_m - a_n| < 1.$$

From this, we obtain with $m_1 = N(1) + 1$ and $n > N(1)$ the bound

$$|a_n| = |a_n - a_{m_1} + a_{m_1}| \leq |a_{m_1} - a_n| + |a_{m_1}| < 1 + |a_{m_1}|.$$

We have, therefore, for all $n \in \mathbb{N}$, the inequality

$$|a_n| \leq \max\{|a_0|, \ldots, |a_{N(1)}|, 1 + |a_{m_1}|\}.$$

This proves the boundedness of the rational Cauchy sequence (a_n).

We consider now the set M of all rational Cauchy sequences, that is,

$$M = \big\{(a_n) \,\big|\, (a_n) \text{ is a rational Cauchy sequence}\big\}.$$

We define on the set M operations of addition and multiplication, which we denote by $+$ and $\cdot$, as follows. For two rational Cauchy sequences (a_n), (b_n), we set

$$(a_n) + (b_n) := (a_n + b_n) \quad \text{and} \quad (a_n) \cdot (b_n) := (a_n \cdot b_n).$$

We must, of course, convince ourselves that the sum and product of two rational Cauchy sequences are again rational Cauchy sequences. This will be proved in the following lemma.

Lemma 2.4. *Let $(a_n), (b_n) \in M$. Then we have*

$$(a_n) + (b_n) \in M \quad \textit{and} \quad (a_n) \cdot (b_n) \in M.$$

Proof. (i) We first prove that the sum $(a_n) + (b_n)$ of the two rational Cauchy sequences (a_n), (b_n) is again a rational Cauchy sequence. To this end, we observe that the sums $a_n + b_n$ are rational for all $n \in \mathbb{N}$. We now choose an arbitrary $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, and note that there exist natural numbers $N_1(\epsilon/2)$ and $N_2(\epsilon/2)$ such that for all $m, n > N := \max\{N_1(\epsilon/2), N_2(\epsilon/2)\}$, we have the inequalities

$$|a_m - a_n| < \frac{\epsilon}{2} \quad \text{and} \quad |b_m - b_n| < \frac{\epsilon}{2}.$$

It now follows from the bound

$$|(a_m + b_m) - (a_n + b_n)| \leq |a_m - a_n| + |b_m - b_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

for $m, n > N$, that $(a_n + b_n)$ is a rational Cauchy sequence, and so, therefore, is the sum $(a_n) + (b_n)$.

(ii) We prove now that the product $(a_n) \cdot (b_n)$ of the two rational Cauchy sequences (a_n), (b_n) is also a rational Cauchy sequence. We note first that the products $a_n \cdot b_n$ are rational for all $n \in \mathbb{N}$. Remark 2.3, on the boundedness of rational Cauchy sequences, allows us to find $c \in \mathbb{Q}$ such that for all $n \in \mathbb{N}$, we have the inequalities

$$|a_n| \leq c \quad \text{and} \quad |b_n| \leq c.$$

Now choose an arbitrary $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, and observe that there exist natural numbers $N_1(\epsilon/(2c))$ and $N_2(\epsilon/(2c))$ such that for all $m, n > N := \max\{N_1(\epsilon/(2c)), N_2(\epsilon/(2c))\}$, we have the inequalities

$$|a_m - a_n| < \frac{\epsilon}{2c} \quad \text{and} \quad |b_m - b_n| < \frac{\epsilon}{2c}.$$

We thereby obtain, for all $m, n > N$, the bound

$$\begin{aligned} |a_m \cdot b_m - a_n \cdot b_n| &= |a_m \cdot b_m - a_m \cdot b_n + a_m \cdot b_n - a_n \cdot b_n| \\ &= |a_m \cdot (b_m - b_n) + b_n \cdot (a_m - a_n)| \\ &\leq |a_m \cdot (b_m - b_n)| + |b_n \cdot (a_m - a_n)| \\ &= |a_m| \cdot |b_m - b_n| + |b_n| \cdot |a_m - a_n| \\ &\leq c \cdot \frac{\epsilon}{2c} + c \cdot \frac{\epsilon}{2c} = \epsilon. \end{aligned}$$

Therefore, $(a_n \cdot b_n)$ is a rational Cauchy sequence, which proves that the product $(a_n) \cdot (b_n)$ is a rational Cauchy sequence. □

Lemma 2.5. *The set of rational Cauchy sequences M together with the additive and multiplicative operations $+$ and $\cdot$, that is, $(M, +, \cdot)$, is a commutative ring with unit element.*

Proof. (i) We show first that $(M, +)$ is an abelian group. We observe first that M is not empty, since it contains the rational Cauchy sequence (0), consisting solely of zeros. The associativity of addition follows at once from the associativity of addition of rational numbers. Namely, if we have $(a_n), (b_n)$, $(c_n) \in M$, then we have also

$$\begin{aligned}\big((a_n)+(b_n)\big)+(c_n) &= (a_n+b_n)+(c_n) = \big((a_n+b_n)+c_n\big)\\ &= \big(a_n+(b_n+c_n)\big) = (a_n)+(b_n+c_n)\\ &= (a_n)+\big((b_n)+(c_n)\big).\end{aligned}$$

The commutativity of addition also follows at once from that of addition of rational numbers. The rational Cauchy sequence (0) mentioned above, consisting solely of zeros, is clearly the additive identity element, since for $(a_n) \in M$, we have

$$(0)+(a_n) = (0+a_n) = (a_n) = (a_n+0) = (a_n)+(0).$$

If $(a_n) \in M$, then we assert that the rational Cauchy sequence $(-a_n)$ is the additive inverse of (a_n). Indeed, we have

$$(-a_n)+(a_n) = (-a_n+a_n) = (0) = (a_n-a_n) = (a_n)+(-a_n).$$

We have thus shown that $(M,+)$ is an abelian group.

(ii) We now show that $(M,\cdot)$ is an abelian monoid. We begin by noting that M is nonempty, since it contains the rational Cauchy sequence (1), consisting solely of ones. The associativity of multiplication of sequences follows at once from the associativity of multiplication of rational numbers. Namely, if $(a_n), (b_n), (c_n) \in M$, then we have

$$\begin{aligned}\big((a_n)\cdot(b_n)\big)\cdot(c_n) &= (a_n\cdot b_n)\cdot(c_n) = \big((a_n\cdot b_n)\cdot c_n\big) = \big(a_n\cdot(b_n\cdot c_n)\big)\\ &= \big(a_n\big)\cdot(b_n\cdot c_n\big) = (a_n)\cdot\big((b_n)\cdot(c_n)\big).\end{aligned}$$

The commutativity of multiplication of sequences follows easily from that of multiplication of rational numbers. The rational Cauchy sequence (1) mentioned above, consisting solely of ones, is clearly the multiplicative identity element in $(M,\cdot)$, since for $(a_n) \in M$, we have

$$(1)\cdot(a_n) = (1\cdot a_n) = (a_n) = (a_n\cdot 1) = (a_n)\cdot(1).$$

We have thus proved that $(M,\cdot)$ is an abelian monoid.

(iii) The validity of the distributive laws for M follows easily from the distributive laws for the rational numbers. For example, for $(a_n), (b_n), (c_n) \in M$, we have

$$\begin{aligned}(a_n)\cdot\big((b_n)+(c_n)\big) &= (a_n)\cdot(b_n+c_n) = \big(a_n\cdot(b_n+c_n)\big)\\ &= (a_n\cdot b_n + a_n\cdot c_n) = (a_n)\cdot(b_n)+(a_n)\cdot(c_n).\end{aligned}$$

The proof of the lemma is thus complete. □

Remark 2.6. If we associate with a rational number r the rational Cauchy sequence (r), the sequence consisting solely of r's, we obtain a mapping

$f : \mathbb{Q} \longrightarrow M$. One may easily check that f is a ring homomorphism

$$f : (\mathbb{Q}, +, \cdot) \longrightarrow (M, +, \cdot).$$

Since clearly, $\ker(f) = \{0\}$, the ring homomorphism f is injective.

Definition 2.7. We set

$$\mathfrak{n} := \big\{(a_n) \in M \,\big|\, (a_n) \text{ is a rational null sequence}\big\}$$

and call this set the *ideal of rational null sequences*. This name is justified by the following lemma.

Lemma 2.8. *The ideal* $\mathfrak{n}$ *of rational null sequences is an ideal in the commutative ring* $(M, +, \cdot)$.

Proof. (i) We must first convince ourselves that $(\mathfrak{n}, +)$ is a subgroup of $(M, +)$. Since the rational Cauchy sequence (0), consisting solely of zeros, is a rational null sequence, it follows that $\mathfrak{n}$ is not empty. Invoking the subgroup criterion, Lemma 2.25 of Chapter II, we see that it suffices to show that for $(a_n), (b_n) \in \mathfrak{n}$, the difference $(a_n) - (b_n)$ is also in $\mathfrak{n}$. Since (a_n) and (b_n) are rational null sequences, there exist for $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, natural numbers $N_a(\epsilon/2)$ and $N_b(\epsilon/2)$ such that for all $n > N_a(\epsilon/2)$ and $n > N_b(\epsilon/2)$, we have the inequalities

$$|a_n| < \frac{\epsilon}{2} \quad \text{and} \quad |b_n| < \frac{\epsilon}{2}.$$

It follows then from the triangle inequality that we have

$$|a_n - b_n| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for $n > \max\{N_a(\epsilon/2), N_b(\epsilon/2)\}$. That is, $(a_n) - (b_n)$ is a rational null sequence. Therefore, $(\mathfrak{n}, +)$ is a subgroup of $(M, +)$.

(ii) Our second task is to show that the product of a rational null sequence $(b_n) \in \mathfrak{n}$ and a rational Cauchy sequence $(a_n) \in M$ is again a rational null sequence. Since the rational Cauchy sequence (a_n) is bounded, by Remark 2.3 (ii), there exists $c \in \mathbb{Q}$, $c > 0$, such that for all $n \in \mathbb{N}$, we have the inequality $|a_n| \leq c$. For an arbitrary $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, there then exists $N(\epsilon/c) \in \mathbb{N}$ such that for all $n > N(\epsilon/c)$, we have the inequality

$$|a_n \cdot b_n| = |a_n| \cdot |b_n| \leq c \cdot \frac{\epsilon}{c} = \epsilon.$$

Therefore, $(a_n) \cdot (b_n)$ is in fact a rational null sequence, and that completes the proof that $\mathfrak{n}$ is an ideal of $(M, +, \cdot)$. □

Remark 2.9. We may now apply Theorem 3.21 of Chapter III to the commutative ring $(M, +, \cdot)$ of rational Cauchy sequences and the ideal $\mathfrak{n}$ of rational null sequences and obtain the commutative quotient ring $(M/\mathfrak{n}, +, \cdot)$. The elements of $M/\mathfrak{n}$ are cosets of the form

$$\alpha = (a_n) + \mathfrak{n},$$

where (a_n) is a rational Cauchy sequence. The elements of each coset are rational Cauchy sequences, and the difference of each pair of elements in the coset is a rational null sequence.

Definition 2.10. Let (a_n) be a rational sequence, and $0 \leq n_0 < n_1 < n_2 < \cdots < n_k < \cdots$ an increasing sequence of natural numbers. The sequence (a_{n_k}) is called a *subsequence* of the sequence (a_n).

Lemma 2.11. *Let (a_n) be a rational Cauchy sequence, and (a_{n_k}) a subsequence of (a_n). Then we have*

$$(a_k) - (a_{n_k}) = (a_k - a_{n_k}) \in \mathfrak{n}.$$

Proof. Let $\epsilon \in \mathbb{Q}$, $\epsilon > 0$. Since (a_n) is a rational Cauchy sequence and $n_k \geq k$, there exists a natural number $N(\epsilon)$ such that for all $k > N(\epsilon)$, we have the inequality

$$|a_k - a_{n_k}| < \epsilon.$$

This shows that the sequence $(a_k - a_{n_k})$ is a rational null sequence, which completes the proof. □

Theorem 2.12. *The quotient ring $(M/\mathfrak{n}, +, \cdot)$ is a field.*

Proof. By construction, the zero and unit elements of $M/\mathfrak{n}$ are given by

$$(0) + \mathfrak{n} \quad \text{and} \quad (1) + \mathfrak{n},$$

where (0) and (1) denote the rational Cauchy sequences that consist solely of zeros and ones respectively.

Since we have already established that $(M/\mathfrak{n}, +, \cdot)$ is a commutative ring with unit element $(1) + \mathfrak{n}$, it remains only to show that every coset $(a_n) + \mathfrak{n}$ other than the zero element of $M/\mathfrak{n}$, that is, for which we have

$$(a_n) + \mathfrak{n} \neq (0) + \mathfrak{n} \quad \Longleftrightarrow \quad (a_n) \notin \mathfrak{n},$$

has a multiplicative inverse. Since $(a_n) \notin \mathfrak{n}$, there exist $\epsilon_0 \in \mathbb{Q}$, $\epsilon_0 > 0$, and $N(\epsilon_0) \in \mathbb{N}$ such that for all $n > N(\epsilon_0)$, we have the inequality

$$|a_n| > \epsilon_0, \quad \text{or equivalently,} \quad a_n \neq 0. \tag{4}$$

We therefore define the rational sequence (b_n) by

$$b_n := \begin{cases} 0, & 0 \leq n \leq N(\epsilon_0), \\ \dfrac{1}{a_n}, & n > N(\epsilon_0). \end{cases}$$

We show first that (b_n) is a rational Cauchy sequence and then that one can construct a multiplicative inverse to $(a_n) + \mathfrak{n}$.

For $m, n > N(\epsilon_0)$, we obtain, using (4),

$$|b_m - b_n| = \left| \frac{1}{a_m} - \frac{1}{a_n} \right| = \frac{|a_m - a_n|}{|a_m \cdot a_n|} < \frac{|a_m - a_n|}{\epsilon_0^2}.$$

Since (a_n) is a rational Cauchy sequence, there exists for $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, a number $N(\epsilon_0^2 \cdot \epsilon)$ such that for all $m, n > N(\epsilon_0^2 \cdot \epsilon)$, we have

$$|a_m - a_n| < \epsilon_0^2 \cdot \epsilon.$$

We thereby obtain at once, for all $m, n > \max\{N(\epsilon_0), N(\epsilon_0^2 \cdot \epsilon)\}$ the inequality

$$|b_m - b_n| < \epsilon;$$

that is, we indeed have $(b_n) \in M$.

Finally, we claim that the element $(b_n) + \mathfrak{n}$ is the multiplicative inverse of $(a_n) + \mathfrak{n}$. To prove this, we have only to show that

$$\big((a_n) + \mathfrak{n}\big) \cdot \big((b_n) + \mathfrak{n}\big) = (1) + \mathfrak{n},$$

that is, that $(a_n) \cdot (b_n) - (1) \in \mathfrak{n}$. For $n > N(\epsilon_0)$, we have by construction that

$$a_n \cdot b_n = 1;$$

that is, the rational Cauchy sequence $(a_n) \cdot (b_n) - (1)$ consists solely of zeros after the first $N(\epsilon_0)$ terms, and it is therefore a rational null sequence. □

Definition 2.13. We call the field $(M/\mathfrak{n}, +, \cdot)$ the *field of real numbers* and denote it by $\mathbb{R}$. In what follows, we shall denote the elements of $\mathbb{R}$ by Greek letters. For example, we have

$$\alpha \in \mathbb{R} \quad \Longleftrightarrow \quad \alpha = (a_n) + \mathfrak{n},$$

for some $(a_n) \in M$.

Lemma 2.14. *The mapping that assigns to every rational number r the real number $(r) + \mathfrak{n}$, where (r) is the rational Cauchy sequence each term of which is equal to r, induces an injective ring homomorphism*

$$F : (\mathbb{Q}, +, \cdot) \longrightarrow (\mathbb{R}, +, \cdot).$$

Proof. For $r_1, r_2 \in \mathbb{Q}$, we verify at once that

$$F(r_1 + r_2) = (r_1 + r_2) + \mathfrak{n} = (r_1 + \mathfrak{n}) + (r_2 + \mathfrak{n}) = F(r_1) + F(r_2),$$
$$F(r_1 \cdot r_2) = (r_1 \cdot r_2) + \mathfrak{n} = (r_1 + \mathfrak{n}) \cdot (r_2 + \mathfrak{n}) = F(r_1) \cdot F(r_2);$$

that is, F is a ring homomorphism. To prove the injectivity of F, we observe that $\ker(F)$ is an ideal of $\mathbb{Q}$. But since $(\mathbb{Q}, +, \cdot)$ is a field, the only ideals of $\mathbb{Q}$ are the zero ideal and the unit ideal (i.e., the entire field). But it is impossible that $\ker(F)$ is the unit ideal, since if it were, every nonzero rational number r would be mapped to the zero element of $\mathbb{R}$, which would imply that the rational Cauchy sequence (r) was a rational null sequence, which is not the case. Therefore, $\ker(F)$ must be the zero ideal, and so F is injective. □

Remark 2.15. From the previous lemma, we may identify the field of rational numbers $\mathbb{Q}$ with its image $\mathrm{im}(F)$ in the field $\mathbb{R}$ of real numbers; that is, we may set $r := (r) + \mathfrak{n}$ $(r \in \mathbb{Q})$.

Definition 2.16. We extend the relations $<$ and $\leq$ on the set $\mathbb{Q}$ of rational numbers from Definition 6.5 of Chapter III to the set $\mathbb{R}$ of real numbers by saying that for two real numbers $\alpha = (a_n) + \mathfrak{n}$ and $\beta = (b_n) + \mathfrak{n}$, we have

$$\alpha < \beta \quad \Longleftrightarrow \quad \exists\, q \in \mathbb{Q}, q > 0, N(q) \in \mathbb{N} : b_n - a_n > q \;\forall n \in \mathbb{N}, n > N(q)$$

and

$$\alpha \leq \beta \quad \Longleftrightarrow \quad \alpha = \beta \text{ or } \alpha < \beta.$$

We extend the definitions of $>$ and $\geq$ similarly to the set $\mathbb{R}$ of real numbers.

Lemma 2.17. *The relation $<$ in Definition 2.16 is well defined, that is, it is independent of the choice of rational Cauchy sequences (a_n) and (b_n) representing the real numbers α and β.*

Proof. We leave the proof as an exercise for the reader. □

Exercise 2.18. Prove Lemma 2.17.

Remark 2.19. With the relation $<$, the set $\mathbb{R}$ of real numbers becomes an *ordered set*; that is, it satisfies the following three conditions:
(i) For every pair of elements $\alpha, \beta \in \mathbb{R}$, one has $\alpha < \beta$ or $\beta < \alpha$ or $\alpha = \beta$.
(ii) The three relations $\alpha < \beta$, $\beta < \alpha$, $\alpha = \beta$, are mutually exclusive.
(iii) If $\alpha < \beta$ and $\beta < \gamma$, then $\alpha < \gamma$.
Analogous statements hold for the relation $>$.

Definition 2.20. Let $\alpha = (a_n) + \mathfrak{n} \in \mathbb{R}$ be a real number. We set

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

We call the real number $|\alpha|$ the *absolute value* of the real number α .

Lemma 2.21. *The absolute value on the real numbers satisfies the following properties:*
(i) $|\alpha \cdot \beta| = |\alpha| \cdot |\beta|$ *for all* $\alpha, \beta \in \mathbb{R}$.
(ii) $|\alpha + \beta| \leq |\alpha| + |\beta|$ *for all* $\alpha, \beta \in \mathbb{R}$.

Proof. We leave the proof as an exercise for the reader. □

Exercise 2.22. Prove Lemma 2.21.

We carry over the notion of a rational Cauchy sequence to the field $(\mathbb{R}, +, \cdot)$ of real numbers.

Definition 2.23. A sequence $(\alpha_n) = (\alpha_n)_{n\geq 0}$ such that $\alpha_n \in \mathbb{R}$ for all $n \in \mathbb{N}$ is called a *real Cauchy sequence* if for every $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, there exists $N(\varepsilon) \in \mathbb{N}$ such that for all $m, n \in \mathbb{N}$ with $m, n > N(\varepsilon)$, we have the inequality

$$|\alpha_m - \alpha_n| < \varepsilon.$$

Remark 2.24. Let (α_n) be a real Cauchy sequence. The nth term of the sequence, α_n, is then given by $\alpha_n = (a_{n,k}) + \mathfrak{n}$, where $(a_{n,k})$ is a rational Cauchy sequence. Moreover, $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, is of the form $\varepsilon = (\epsilon_k) + \mathfrak{n}$ with the rational Cauchy sequence (ϵ_k). Using Definition 2.16 with the relation $<$, Definition 2.23 now takes the form that for $m, n \in \mathbb{N}$ with $m, n > N(\varepsilon)$, there exists a natural number $M(m, n)$ such that for all $k > M(m, n)$, we have the inequality

$$|a_{m,k} - a_{n,k}| < \epsilon_k.$$

Exercise 2.25. Give examples of real null sequences whose terms are all irrational (that is, not rational) numbers.

Definition 2.26. A real sequence (α_n) has a *limit* $\alpha \in \mathbb{R}$, or equivalently, the sequence *converges* to $\alpha \in \mathbb{R}$, if for every $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, there exists a natural number $N(\varepsilon)$ such that for all $n \in \mathbb{N}$ with $n > N(\varepsilon)$, we have the inequality

$$|\alpha_n - \alpha| < \varepsilon.$$

In this case, we write

$$\alpha = \lim_{n\to\infty} \alpha_n.$$

Theorem 2.27. *In the field* $(\mathbb{R}, +, \cdot)$ *of real numbers, every real Cauchy sequence* (α_n) *has a limit* $\alpha \in \mathbb{R}$.

Proof. By Remark 2.24, we have $\alpha_n = (a_{n,k}) + \mathfrak{n}$ with the rational Cauchy sequence $(a_{n,k})$. We shall show that

(i) the rational sequence $(a_{n,n})$ is a Cauchy sequence;
(ii) we have $\lim_{n\to\infty} \alpha_n = \alpha$, where $\alpha := (a_{n,n}) + \mathfrak{n}$.

(i) Let $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$; without loss of generality, we may choose ε to be rational; that is, we may have $\varepsilon = (\epsilon)$ with $\epsilon \in \mathbb{Q}$. By Remark 2.24, there exists for all $m, n > N(\epsilon)$, a natural number $M(m,n)$ such that for all $k > M(m,n)$, we have the inequality

$$|a_{m,k} - a_{n,k}| < \epsilon.$$

We now show that there exists a natural number $N_0(\epsilon)$ such that the inequality

$$|a_{m,k} - a_{n,k}| < \epsilon$$

holds for all $m, n > N_0(\epsilon)$ and all $k \in \mathbb{N}$. To this end, we observe first of all that the rational Cauchy sequence $(a_{n,k})$ representing α_n can be altered by passage to a subsequence, which for simplicity we again denote by $(a_{n,k})$, in such a way that

$$|a_{n,k} - a_{n,n}| < \frac{1}{n} \tag{5}$$

for all $k \in \mathbb{N}$. It then follows by the triangle inequality that for arbitrary $k, k' \in \mathbb{N}$, we have

$$|a_{n,k} - a_{n,k'}| \leq |a_{n,k} - a_{n,n}| + |a_{n,n} - a_{n,k'}| < \frac{2}{n}.$$

We thereby obtain for all $m, n > N(\epsilon/2)$, $k \in \mathbb{N}$, and $k' > M(m,n)$ the bound

$$\begin{aligned} |a_{m,k} - a_{n,k}| &\leq |a_{m,k} - a_{m,k'}| + |a_{m,k'} - a_{n,k'}| + |a_{n,k'} - a_{n,k}| \\ &< \frac{2}{m} + \frac{\epsilon}{2} + \frac{2}{n}. \end{aligned}$$

If we now set $N_0(\epsilon) := \max\{N(\frac{\epsilon}{2}), [\frac{8}{\epsilon}]\}$, we obtain, as desired, for all $m, n > N_0(\epsilon)$ and all $k \in \mathbb{N}$, the bound

$$|a_{m,k} - a_{n,k}| < \epsilon. \tag{6}$$

If we now choose $m, n > N_0(\epsilon)$ and set $k = m$ in the inequalities (6) and (5), we obtain

$$\begin{aligned} |a_{m,m} - a_{n,n}| &\leq |a_{m,m} - a_{n,m}| + |a_{n,m} - a_{n,n}| \\ &< \epsilon + \frac{1}{n} < \epsilon + \frac{\epsilon}{8} < 2\epsilon. \end{aligned}$$

We have thus shown that $(a_{n,n})$ is a rational Cauchy sequence by which we define the real number α; that is, we have $\alpha := (a_{n,n}) + \mathfrak{n}$.

(ii) It remains to show that $\lim_{n\to\infty} \alpha_n = \alpha$. Again let $\epsilon \in \mathbb{Q}$, $\epsilon > 0$. We must show that for sufficiently large n, k, we have the inequality

$$|a_{n,k} - a_{k,k}| < \epsilon. \tag{7}$$

Setting $m = k$ into (6) and then choosing $n, k > N_0(\epsilon)$, we obtain the desired inequality (7), that is,

$$|\alpha_n - \alpha| < \epsilon$$

for all $n > N_0(\epsilon)$. This proves the assertion. □

Definition 2.28. Because of the fact proved in the previous theorem that every real Cauchy sequence in the field $(\mathbb{R}, +, \cdot)$ has a limit that itself is in $\mathbb{R}$, we say that the real numbers are *complete*.

Remark 2.29. Let $\alpha = (a_n) + \mathfrak{n} \in \mathbb{R}$ be a real number. The rational Cauchy sequence (a_n) is in particular also a real Cauchy sequence, which we may also denote by (a_n) because we have identified $\mathbb{Q}$ with a subset of the real numbers $\mathbb{R}$. The proof of Theorem 2.27 shows that

$$\alpha = \lim_{n\to\infty} a_n;$$

that is, every real number is the limit of a rational Cauchy sequence.

Exercise 2.30. Find a rational Cauchy sequence with limit $\sqrt{2}$.

3. The Decimal Expansion of a Real Number

Definition 3.1. Let q_{-j} be natural numbers with $0 \le q_{-j} \le 9$ for

$$j = -\ell, \ldots, 0, 1, 2, \ldots$$

and $\ell \in \mathbb{N}$. Then we call the formal sum

$$\pm q_\ell \ldots q_0 . q_{-1} q_{-2} \ldots := \pm \sum_{j=-\ell}^{\infty} q_{-j} \cdot 10^{-j}$$

an *(infinite) decimal*. We set

$$\mathbb{D}' := \{\pm q_\ell \ldots q_0 . q_{-1} q_{-2} \ldots \mid \pm q_\ell \ldots q_0 . q_{-1} q_{-2} \ldots \text{ is a decimal}\}.$$

The decimal $\pm q_\ell \ldots q_0 . q_{-1} q_{-2} \ldots$ is said to be *terminating* if there exists an index $k \ge 0$ such that $q_{-j} = 0$ for $j > k$. The notion of periodicity of a decimal expansion and the associated notions from Definition 1.3 can be carried over to decimals without further ado.

Remark 3.2. On the basis of our previous considerations, terminating and periodic decimals can be identified with rational numbers. The remaining decimals have no clear meaning at this point. However, the following

lemma will allow us to identify them with real numbers. To this end, we make an association between the set $\mathbb{D}'$ of decimals and the set $\mathbb{R}$ of real numbers.

Lemma 3.3. *Let $\pm q_\ell \dots q_0.q_{-1}q_{-2}\dots$ be a decimal and (a_n) the rational sequence given by*

$$a_n := \pm q_\ell \dots q_0.q_{-1}\dots q_{-n}.$$

With the assignment

$$\pm q_\ell \dots q_0.q_{-1}q_{-2}\dots \;\mapsto\; (a_n) + \mathfrak{n},$$

we obtain a surjective *mapping of sets*

$$\varphi : \mathbb{D}' \longrightarrow \mathbb{R}.$$

Proof. (i) We must first demonstrate that the mapping φ is well defined, that is, that the sequence (a_n) is a rational Cauchy sequence. To this end, let $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, and $N \in \mathbb{N}$ be such that $10^{-N} < \epsilon$. Then by construction, we have that for all $m, n > N$ with $m > n$,

$$|a_m - a_n| < 0.0\dots0\,q_{-(n+1)}\dots q_{-m} < 10^{-N} < \epsilon,$$

which proves the Cauchy sequence property.

(ii) We now prove the surjectivity of φ. It suffices to show that φ yields a surjective mapping of the set of all nonnegative decimals

$$q_\ell \dots q_0.q_{-1}q_{-2}\dots \in \mathbb{D}'$$

to the set of nonnegative real numbers. To construct a decimal expansion of a given real number, we shall try to imitate the procedure given in Section 1 for obtaining the decimal expansion of a rational number. However, we do not have, in general, division with remainder at our disposal. As a substitute, we shall use the fact that we can decompose a real number into its integer part and fractional part, the latter being nonnegative and less than 1.

Suppose, then, that we have $\alpha \in \mathbb{R}$ with $\alpha \geq 0$. There then exist $q \in \mathbb{N}$ and $\rho \in \mathbb{R}$ with $0 \leq \rho < 1$ such that

$$\alpha = q + \rho.$$

For the natural number q, we have the decimal representation

$$q = \sum_{j=0}^{\ell} q_j \cdot 10^j = q_\ell q_{\ell-1} \dots q_1 q_0.$$

We write

$$\rho = \frac{1}{10} \cdot 10 \cdot \rho$$

and decompose $10 \cdot \rho$ as before in the form

$$10 \cdot \rho = q_{-1} + \rho_{-1}$$

with $q_{-1} \in \mathbb{N}$ and $\rho_{-1} \in \mathbb{R}$ with $0 \leq \rho_{-1} < 1$; since we have $\rho < 1$, it follows that $0 \leq q_{-1} \leq 9$. We thereby obtain

$$\rho = \frac{1}{10}(q_{-1} + \rho_{-1}) = \frac{q_{-1}}{10} + \frac{\rho_{-1}}{10}.$$

We again write

$$10 \cdot \rho_{-1} = q_{-2} + \rho_{-2},$$

where $q_{-2} \in \mathbb{N}$ with $0 \leq q_{-2} \leq 9$ and $\rho_{-2} \in \mathbb{R}$ with $0 \leq \rho_{-2} < 1$. This yields

$$\rho = \frac{q_{-1}}{10} + \frac{1}{10^2}(q_{-2} + \rho_{-2}) = \frac{q_{-1}}{10} + \frac{q_{-2}}{10^2} + \frac{\rho_{-2}}{10^2}.$$

Proceeding in this way, we obtain the decimal expansion

$$q_\ell \dots q_0 . q_{-1} q_{-2} \dots \in \mathbb{D}'.$$

The partial sums

$$a_n = q_\ell \dots q_0 . q_{-1} \dots q_{-n}$$

of this decimal expansion form a rational Cauchy sequence, which by construction, converges in $\mathbb{R}$ to α. We thereby see that

$$\varphi(q_\ell \dots q_0 . q_{-1} q_{-2} \dots) = \alpha.$$

This proves the surjectivity of φ. □

Remark 3.4. The proof of the preceding Lemma 3.3 shows that the decimal $\pm q_\ell \dots q_0 . q_{-1} q_{-2} \dots$ corresponds to the real number

$$\alpha = \pm \lim_{n \to \infty} \sum_{j=-\ell}^{n} q_{-j} \cdot 10^{-j} = \pm \sum_{j=-\ell}^{\infty} q_{-j} \cdot 10^{-j}$$

via the mapping φ. We have thus answered in the affirmative the question raised in Section 1 of this chapter about the possible meaning of such series.

Remark 3.5. We now investigate the mapping $\varphi : \mathbb{D}' \longrightarrow \mathbb{R}$ from Lemma 3.3 with respect to injectivity. We shall see that φ is not injective. It will therefore be our goal to measure the defect in injectivity. In what follows, we may again restrict our attention to the set of nonnegative decimals.

Lemma 3.6. *Let*

$$q_\ell \dots q_0 . q_{-1} q_{-2} \dots \quad \text{and} \quad q'_{\ell'} \dots q'_0 . q'_{-1} q'_{-2} \dots$$

be nonnegative decimals such that

$$\varphi(q_\ell \dots q_0 . q_{-1} q_{-2} \dots) = \varphi(q'_{\ell'} \dots q'_0 . q'_{-1} q'_{-2} \dots), \tag{8}$$

where φ is the set mapping from $\mathbb{D}'$ to $\mathbb{R}$ defined in Lemma 3.3. Then either the two decimals are identical, or one of the two terminates and the other consists solely of 9's from some point in its decimal expansion onward.

Proof. We may, without loss of generality, assume $\ell' \geq \ell$. By Remark 3.4, we obtain from (8) the equality

$$q'_{\ell'} \cdot 10^{\ell'} + \dots + q'_{\ell+1} \cdot 10^{\ell+1} = \sum_{j=-\ell}^{\infty} (q_{-j} - q'_{-j}) \cdot 10^{-j}.$$

If we divide this equation by $10^{\ell+1}$, we shall see that without loss of generality, we may assume that $\ell = -1$. We thus obtain

$$q'_{\ell'} \cdot 10^{\ell'} + \dots + q'_0 = \sum_{j=1}^{\infty} (q_{-j} - q'_{-j}) \cdot 10^{-j}. \tag{9}$$

Since $0 \leq q_j, q'_j \leq 9$, we may estimate the right-hand side of (9), obtaining

$$\begin{aligned} 0 \leq \left| \sum_{j=1}^{\infty} (q_{-j} - q'_{-j}) \cdot 10^{-j} \right| &\leq \sum_{j=1}^{\infty} |q_{-j} - q'_{-j}| \cdot 10^{-j} \leq 9 \cdot \sum_{j=1}^{\infty} 10^{-j} \\ = 9\left(\sum_{j=0}^{\infty} 10^{-j} - 1 \right) &= 9\left(\frac{1}{1 - \frac{1}{10}} - 1 \right) = 9\left(\frac{10}{9} - 1 \right) = 1. \end{aligned}$$

We therefore obtain

$$0 \leq \left(q'_{\ell'} \cdot 10^{\ell'} + \dots + q'_0 \right) \leq 1$$

for the left-hand side of (9); that is, we have $\ell' = 0$ and $q'_0 = 1$ or $\ell' = -1$, whence $q'_0 = 0$. Since the first case obtains precisely when equality holds in the previous inequality for all $j = 1, 2, \dots$, we see that this case obtains only if for $j = 1, 2, \dots$, we have the equality

$$|q_{-j} - q'_{-j}| = 9.$$

Since this is all taking place with nonnegative numbers, this means that

$$q_{-j} = 9 \quad \text{and} \quad q'_{-j} = 0 \quad (j = 1, 2, \dots).$$

If the latter case obtains, then we proceed from the equality $q_0 = q'_0$ and look for an index $-k$ such that $q_{-k} = q'_{-k}$, yet $q_{-k-1} \neq q'_{-k-1}$. Either there is no such index, in which case the two decimals are identical, or there indeed

exists such an index $-k$; if we then argue as above, we see that the decimal $0, q_{-1}q_{-2}\ldots$ consists solely of 9's from the $(k+1)$st decimal place onward. □

Definition 3.7. We now define $\mathbb{D} \subset \mathbb{D}'$ as the subset containing no decimals containing only 9's from some point on. We shall call the elements of $\mathbb{D}$ *genuine decimals*.

Theorem 3.8. *There is a bijection between the set $\mathbb{D}$ of genuine decimals and the set $\mathbb{R}$ of real numbers.*

Proof. The theorem follows directly from Lemmas 3.3 and 3.6. □

Remark 3.9. Using Theorem 3.8, we may henceforward speak of the *decimal representation* or *decimal expansion* of real numbers.

Moreover, with the help of the bijection between $\mathbb{D}$ and $\mathbb{R}$ from Theorem 3.8, we may carry over addition and multiplication of real numbers to the set of genuine decimals. We thereby obtain the field $(\mathbb{D}, +, \cdot)$ of genuine decimals.

Remark 3.10. Using the decimal representation of real numbers, it can be shown that the set $\mathbb{R}$ is uncountable. We are not going to discuss this further, and we refer the reader to the relevant literature. Since the set $\mathbb{Q}$ of rational numbers is countable, the set difference $\mathbb{R} \setminus \mathbb{Q}$ is not empty. This fact leads to the following definition.

Definition 3.11. A real number $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is said to be *irrational*.

Exercise 3.12.
(a) Think about why the number $0.101\,001\,000\,100\,001\ldots$ (that is, the number of zeros between ones is successively $1,2,3,\ldots$) is irrational. Give further examples of irrational decimals.
(b) Compute the first ten decimal places of $\sqrt{2}$ precisely.

4. Equivalent Characterizations of Completeness

In this section, we shall present some equivalent characterizations of the completeness of the real numbers. This will lead us to the notions of supremum and infimum.

Definition 4.1. A real sequence (α_n) is said to be *monotonically increasing* if for all $n \in \mathbb{N}$, we have the inequality $\alpha_{n+1} \geq \alpha_n$, and it is said to be *strictly monotonically increasing* if we have $\alpha_{n+1} > \alpha_n$.

A real sequence (α_n) is said to be *monotonically decreasing* if for all $n \in \mathbb{N}$, we have the inequality $\alpha_{n+1} \leq \alpha_n$, and it is said to be *strictly monotonically decreasing* if we have $\alpha_{n+1} < \alpha_n$.

Exercise 4.2. Determine which of the following sequences are (strictly) monotonically increasing or (strictly) monotonically decreasing:

$$\left(12^{\frac{1}{n+1}}\right)_{n\geq 0}, \quad \left(\frac{n^3-2}{n^2-2}\right)_{n\geq 0}, \quad \left(\frac{n^2+2}{2^n}\right)_{n\geq 0}, \quad \left(\frac{n^3+3}{3^n}\right)_{n\geq 0}, \quad \left(n^{\frac{1}{n+1}}\right)_{n\geq 0}.$$

Definition 4.3. A nonempty set $\mathfrak{M} \subseteq \mathbb{R}$ is said to be *bounded from above* if there exists $\gamma \in \mathbb{R}$ such that for all $\mu \in \mathfrak{M}$, we have $\mu \leq \gamma$. The real number γ is called an *upper bound* for the set $\mathfrak{M}$.

A nonempty set $\mathfrak{M} \subseteq \mathbb{R}$ is said to be *bounded from below* if there exists $\gamma \in \mathbb{R}$ such that for all $\mu \in \mathfrak{M}$, we have $\mu \geq \gamma$. The real number γ is called a *lower bound* for the set $\mathfrak{M}$.

A nonempty set $\mathfrak{M} \subseteq \mathbb{R}$ is said to be *bounded* if it is bounded both from above and from below.

Theorem 4.4. *If a nonempty set $\mathfrak{M} \subseteq \mathbb{R}$ is bounded from above, then the set $\mathfrak{M}$ has a least upper bound $\sigma \in \mathbb{R}$.*

Proof. We choose $\alpha_0 \in \mathbb{R}$ such that α_0 is not an upper bound for $\mathfrak{M}$ but such that $\beta_0 := \alpha_0 + 1$ is such an upper bound. Then $\alpha_0 + \frac{1}{2}$ is either an upper bound for $\mathfrak{M}$ or it is not. In the first case, we define

$$\alpha_1 := \alpha_0 \quad \text{and} \quad \beta_1 := \alpha_0 + \frac{1}{2},$$

while in the second case, we define

$$\alpha_1 := \alpha_0 + \frac{1}{2} \quad \text{and} \quad \beta_1 := \beta_0.$$

Proceeding in this fashion, we construct inductively two real sequences (α_n) and (β_n) that satisfy the following three properties:
(1) The sequence (α_n) is monotonically increasing.
(2) The sequence (β_n) is monotonically decreasing.
(3) For all $m, n \in \mathbb{N}$, we have the inequality $\alpha_n \leq \beta_m$.

Choose now $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, and $m \in \mathbb{N}$ such that $2^{-m} < \varepsilon$. Then for all $n \in \mathbb{N}$ with $n > m$, we have by the nature of our construction that

$$|\alpha_m - \alpha_n| = \alpha_n - \alpha_m \leq \beta_m - \alpha_m \leq \frac{1}{2^m} < \varepsilon.$$

We see, then, that (α_n) is a real Cauchy sequence. Analogously, we see that (β_n) is also a real Cauchy sequence. We set

$$\alpha := \lim_{n\to\infty} \alpha_n, \quad \beta := \lim_{n\to\infty} \beta_n.$$

Since

$$\lim_{n\to\infty} (\beta_n - \alpha_n) = 0,$$

we conclude that $\alpha = \beta$.

We now claim that α is the desired least upper bound for the set $\mathfrak{M}$. Since the real numbers β_n are by construction upper bounds for $\mathfrak{M}$ for all $n \in \mathbb{N}$, we have for all $\mu \in \mathfrak{M}$ and $n \in \mathbb{N}$ the inequality

$$\mu \leq \beta_n,$$

that is,

$$\mu \leq \beta = \alpha.$$

Thus α is an upper bound for $\mathfrak{M}$.

Now let $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, be such that $\alpha' := \alpha - \varepsilon$ is a smaller upper bound for $\mathfrak{M}$. By the monotonicity of the sequence (α_n), we can find $N(\varepsilon) \in \mathbb{N}$ such that for all $n > N(\varepsilon)$, we have the inequality

$$\alpha_n \geq \alpha - \varepsilon = \alpha'.$$

But since α_n by construction cannot be an upper bound for $\mathfrak{M}$ for all $n \in \mathbb{N}$, there must exist $\mu_n \in \mathfrak{M}$ such that $\mu_n > \alpha_n$. If we now choose $n > N(\varepsilon)$, we obtain the contradiction $\mu_n > \alpha'$. Therefore, α is the least upper bound for $\mathfrak{M}$, and the theorem is proved. □

One can prove analogously the following theorem.

Theorem 4.5. *If a nonempty set $\mathfrak{M} \subseteq \mathbb{R}$ is bounded from below, then the set $\mathfrak{M}$ has a greatest lower bound $\sigma \in \mathbb{R}$.* □

Exercise 4.6. Give an example showing that there is no valid analogue of Theorem 4.5 for the set of rational numbers.

Exercise 4.7. Find the greatest lower bound and the least upper bound for the set $\{\sqrt[x]{x} \mid x \in \mathbb{Q},\ x \geq 0\}$.

Definition 4.8. The least upper bound for a nonempty set $\mathfrak{M}$ that is bounded above, as described above in Theorem 4.4, is called the *supremum* of $\mathfrak{M}$ and is denoted by $\sup(\mathfrak{M})$.

The greatest lower bound for a nonempty set $\mathfrak{M}$ that is bounded below is called the *infimum* of $\mathfrak{M}$ and is denoted by $\inf(\mathfrak{M})$.

Definition 4.9. A sequence of closed intervals

$$[\alpha_n, \beta_n] := \{\delta \in \mathbb{R} \mid \alpha_n \leq \delta \leq \beta_n\} \subseteq \mathbb{R} \quad (n \in \mathbb{N})$$

is called a sequence of *nested intervals* if the real sequences (α_n) and (β_n) satisfy the following three properties:
(1) The sequence (α_n) is monotonically increasing.
(2) The sequence (β_n) is monotonically decreasing.
(3) We have the limit $\lim_{n\to\infty}(\beta_n - \alpha_n) = 0$.

Theorem 4.10. *If the intervals $[\alpha_n, \beta_n] \subseteq \mathbb{R}$ for $n \in \mathbb{N}$ form a sequence of nested intervals, then we have*

$$\bigcap_{n=0}^{\infty} [\alpha_n, \beta_n] = \{\alpha\}$$

for some real number α.

Proof. We begin by showing that the intersection

$$\bigcap_{n=0}^{\infty} [\alpha_n, \beta_n]$$

is nonempty. To this end, we consider the nonempty set

$$\mathfrak{A} := \{\alpha_n \mid n \in \mathbb{N}\} \subseteq \mathbb{R} \quad \text{and} \quad \mathfrak{B} := \{\beta_n \mid n \in \mathbb{N}\} \subseteq \mathbb{R}.$$

By definition, the set $\mathfrak{A}$ is bounded from above, namely by the elements of $\mathfrak{B}$. Similarly, the set $\mathfrak{B}$ is bounded from below. By Theorems 4.4 and 4.5, we may consider the supremum of $\mathfrak{A}$ and the infimum of $\mathfrak{B}$, that is,

$$\alpha := \sup(\mathfrak{A}) \in \mathbb{R} \quad \text{and} \quad \beta := \inf(\mathfrak{B}) \in \mathbb{R}.$$

By property (3) in Definition 4.9, we must have $\alpha = \beta$. Since $\alpha_n \leq \alpha = \beta \leq \beta_n$ for all $n \in \mathbb{N}$, we have found with α an element that belongs to every interval $[\alpha_n, \beta_n]$ $(n \in \mathbb{N})$.

We now show that α is the only element in the intersection under discussion. To this end, let $\gamma \in \bigcap_{n=0}^{\infty}[\alpha_n, \beta_n]$ be arbitrary. Since $\alpha_n \leq \gamma \leq \beta_n$ for all $n \in \mathbb{N}$, we have

$$\alpha = \lim_{n\to\infty} \alpha_n \leq \gamma \leq \lim_{n\to\infty} \beta_n = \beta,$$

which shows that $\alpha = \gamma$. □

Remark 4.11. We may paraphrase the result of Theorem 4.10 by saying that the *nested intervals principle* holds in $\mathbb{R}$. We have seen that the completeness of the real numbers, called the *completeness principle*, has as a consequence the existence of a supremum (*supremum principle*) and an infimum (*infimum principle*), which in turn implies the nested intervals principle. We close this circle by showing that the nested intervals principle implies completeness. In sum, in the set of real numbers, the following are equivalent:
– the completeness principle,

– the supremum and infimum principles,
– the nested intervals principle.

Theorem 4.12. *Consider the set $(\mathbb{R}, +, \cdot)$ of real numbers with its order relation $<$ and assume the validity of the nested intervals principle. Then every real Cauchy sequence (α_n) has a limit in $\mathbb{R}$; that is, the nested intervals principle implies the completeness principle.*

Proof. Let $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$. Then there exists $N(\varepsilon) \in \mathbb{N}$ such that for all natural numbers $m, n > N(\varepsilon)$, we have the inequality

$$|\alpha_m - \alpha_n| < \varepsilon.$$

If $n_0 := N(\varepsilon) + 1$, then for all natural numbers $n \geq n_0$, we have

$$|\alpha_n - \alpha_{n_0}| < \varepsilon,$$

that is,

$$|\alpha_n| < |\alpha_{n_0}| + \varepsilon.$$

Setting

$$\mu := \max\{|\alpha_0|, \ldots, |\alpha_{n_0-1}|, |\alpha_{n_0}| + \varepsilon\},$$

we see that $\mathfrak{M} := \{\alpha_n \mid n \in \mathbb{N}\} \subseteq \mathbb{R}$ is a bounded set; that is, there exist real numbers μ_0, ν_0 such that for all $n \in \mathbb{N}$, we have the inequalities

$$\mu_0 \leq \alpha_n \leq \nu_0.$$

By repeated halving of the closed interval $[\mu_0, \nu_0]$, we obtain a sequence of nested intervals $[\mu_k, \nu_k]$ $(k \in \mathbb{N})$ such that infinitely many terms of the sequence lie in each of the intervals; that is, for infinitely many indices n, we have

$$\mu_k \leq \alpha_n \leq \nu_k.$$

The assumed validity of the nested intervals principle implies the existence of a real number α that is determined by

$$\bigcap_{k=0}^{\infty} [\mu_k, \nu_k] = \{\alpha\}.$$

There exists, therefore, a natural number $K(\varepsilon)$ such that for all natural numbers $k > K(\varepsilon)$, we have the inequalities

$$\alpha - \varepsilon < \mu_k < \nu_k < \alpha + \varepsilon;$$

that is, for infinitely many indices m, we have the inequalities

$$\alpha - \varepsilon < \alpha_m < \alpha + \varepsilon \quad \Longleftrightarrow \quad |\alpha_m - \alpha| < \varepsilon.$$

Increasing $N(\varepsilon)$ if necessary, we may choose one of the infinitely many indices $m = n_0$ and obtain

$$|\alpha_{n_0} - \alpha| < \varepsilon.$$

It therefore follows that for all natural numbers $n > N(\varepsilon)$, we have

$$|\alpha_n - \alpha| \leq |\alpha_n - \alpha_{n_0}| + |\alpha_{n_0} - \alpha| < 2\varepsilon.$$

This proves the convergence of the real Cauchy sequence (α_n) and that the equality

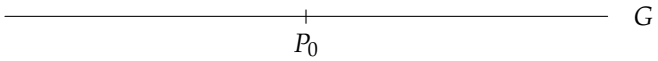
$$\lim_{n\to\infty} \alpha_n = \alpha$$

holds. □

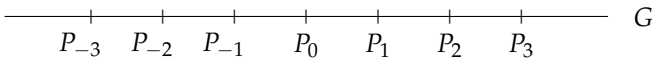
5. The Real Numbers and the Real Number Line

In this section, we shall construct a bijection between the elements of the set of real numbers and the points on a straight line. This will lead us to the notion of the *real number line*. To do so, we shall require the classical axioms of plane Euclidean geometry. We shall use in particular that the plane consists of points, that through every two points in the plane there passes exactly one straight line and this line determines a unique line segment (connecting the two points), and that two distinct lines in the plane have exactly one point of intersection unless they are parallel. We shall also require the fact that we can mark off segments on a line using a compass. We shall also assume the validity of the important similarity theorems. We shall see, however, that the classical axioms do not suffice for carrying out our desired identification of the set of real numbers with a straight line. We shall need an additional axiom, which we shall call the *axiom of geometric completeness*.

We begin with the set $\mathbb{R}$ of real numbers and a horizontal straight line G in the plane. Our goal is to produce a bijection from the set $\mathbb{R}$ of real numbers to the line G. We begin by choosing a point P_0 on the line G, which we call the zero point. We choose the point $P_0 \in G$ as the image of the zero element $0 \in \mathbb{R}$.



After marking off an arbitrary but fixed unit distance on the line, we can mark equidistant points on the line G, beginning from the zero point P_0 and moving to the right, a unit distance apart, which we consider the images of the natural numbers $1, 2, 3, \ldots$. We denote these points by $P_1, P_2, P_3, \ldots$. By reflecting in the zero point P_0, we obtain the images of the negative integers $-1, -2, -3, \ldots$ on G, which we denote by $P_{-1}, P_{-2}, P_{-3}, \ldots$.



If we define the length $\ell(\overline{P_0P_1})$ of the unit line segment $\overline{P_0P_1}$ to be 1, then the length of the segment $\overline{P_aP_b}$ ($a, b \in \mathbb{Z}$, $a \leq b$) will be

$$\ell(\overline{P_aP_b}) = b - a.$$

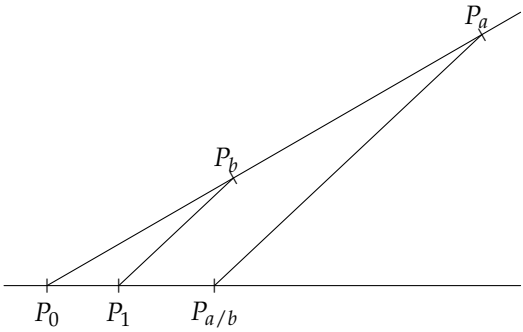
We now imagine two intersecting lines in the plane on which the integer points have been marked and that intersect at the zero point P_0. On one line, we mark the points P_a and P_b corresponding to the natural numbers a and b ($a, b \neq 0$), and on the other line, we mark the point P_1, corresponding to the natural number 1. If we now construct the segment joining points P_b and P_1 and then construct the line parallel to this segment through the point P_a, we obtain as the intersection point of this parallel line with the other straight line a point P. By the intercept theorem, we have the following relationship between the segments $\overline{P_0P_b}$, $\overline{P_0P_a}$, $\overline{P_0P_1}$, and $\overline{P_0P}$:

$$\overline{P_0P_b} : \overline{P_0P_a} = \overline{P_0P_1} : \overline{P_0P}.$$

If we denote the length of the segment $\overline{P_0P}$ by x, we have the following proportion between the lengths of the segments under consideration:

$$b : a = 1 : x \quad \Longleftrightarrow \quad a = b \cdot x \quad \Longleftrightarrow \quad x = \frac{a}{b}.$$

We may therefore consider the point P to be the image of the positive rational number $\frac{a}{b}$, and we denote it by $P_{a/b}$.



By carrying out this construction for all positive rational numbers, we obtain for every rational number r an image point $P_r \in G$. Again, by reflection in the zero point, we obtain as well the negative rational numbers as points on G. In sum, we have obtained an injective mapping $\psi : \mathbb{Q} \longrightarrow G$ through the assignment $r \mapsto P_r$.

Before we extend the mapping ψ to the set of real numbers, let us note some important properties of ψ that result from how it was constructed.

First, ψ preserves the order relation $<$ on $\mathbb{Q}$ in the sense that for rational numbers r, s with $r < s$, the image point P_r lies to the *left* of the image point P_s. Moreover, the mapping ψ respects the addition and multiplication of rational numbers. For example, for two positive rational numbers r, s, we obtain the image point P_{r+s} of the sum $r + s$ as the point of G obtained as the concatenation of the segments $\overline{P_0P_r}$ and $\overline{P_0P_s}$. The difference of two rational numbers is interpreted similarly. The point $P_{r\cdot s}$ corresponding to the product $r \cdot s$ of two (positive) rational numbers r, s can be constructed by a suitable application of the intercept theorem to the segments $\overline{P_0P_r}$, $\overline{P_0P_s}$.

We are ready now to extend the mapping $\psi : \mathbb{Q} \longrightarrow G$ to the set $\mathbb{R}$ of real numbers. For simplicity of notation, we shall continue to denote this extended mapping by ψ. We begin by defining the image of an interval $[r, s] \subseteq \mathbb{R}$ with *rational* endpoints r, s as the set of points on the segment $\overline{P_rP_s} \subseteq G$, that is,

$$\psi([r, s]) = \overline{P_rP_s}.$$

We note that the length of the segment $\overline{P_rP_s}$ is given by

$$\ell(\overline{P_rP_s}) = s - r.$$

Now let $\alpha \in \mathbb{R}$ be an arbitrary real number. The construction of the real numbers together with the nested intervals principle shows that we can obtain α as the intersection of the intervals $I_n = [a_n, b_n]$ $(n \in \mathbb{N})$ of a sequence of nested intervals with *rational* endpoints. The intervals $I_n = [a_n, b_n]$ $(n \in \mathbb{N})$ are mapped by ψ to the segments $\overline{P_{a_n}P_{b_n}}$. To continue the mapping ψ from $\mathbb{Q}$ to $\mathbb{R}$, we require at this point an additional axiom.

Definition 5.1. The line G satisfies the *axiom of geometric completeness* if every sequence of nested line segments $\overline{P_{a_n}P_{b_n}}$ $(n \in \mathbb{N})$ with

$$\lim_{n\to\infty} \ell\left(\overline{P_{a_n}P_{b_n}}\right) = 0$$

has a nonempty intersection, that is, if we have

$$\bigcap_{n=0}^{\infty} \overline{P_{a_n}P_{b_n}} \neq \varnothing. \tag{10}$$

Remark 5.2. We see that the intersection (10) in Definition 5.1 consists of a single point $P \in G$: First of all, the intersection (10) is nonempty by the axiom of geometric completeness. For the sake of obtaining a contradiction, let us suppose that the intersection (10) contains at least the two points P, Q. These two points are separated by some positive distance $d > 0$. If we choose n large enough, then the length of the segment $\overline{P_{a_n}P_{b_n}}$ will be less than d. This leads to a contradiction, since under these conditions, the points P, Q cannot both belong to the intersection (10).

Using the axiom of geometric completeness and taking into account the above observation, we now set

$$\psi(\alpha) := \bigcap_{n=0}^{\infty} \overline{P_{a_n} P_{b_n}} = \{P\}.$$

We see easily that this definition is independent of the choice of sequence of nested intervals. We thereby obtain a mapping ψ from the set $\mathbb{R}$ of real numbers to the set of points on the line G. As in the case of the rational numbers, we have that the mapping ψ respects the order relation $<$. We say that the mapping ψ is *order-preserving*. This gives us at once the injectivity of ψ.

Finally, we observe that the mapping ψ is also surjective. To see this, let $P \in G$ be a point. We consider the set

$$\mathfrak{M} := \{r \in \mathbb{Q} \mid \psi(r) \text{ lies to the left of } P\}.$$

Since there exists a natural number a whose image point P_a lies to the right of P, the nonempty set $\mathfrak{M}$ is bounded from above. There exists, therefore, the supremum for $\mathfrak{M}$; we set $\alpha := \sup(\mathfrak{M}) \in \mathbb{R}$ and claim that $P = \psi(\alpha)$. If we had $P \neq \psi(\alpha)$, then we would have $\psi(\alpha)$ strictly to the left of P, due to the order-preserving nature of ψ; that is, the segment from $\psi(\alpha)$ to P would have positive length. If we now choose a monotonically decreasing sequence (c_n) of rational numbers that converges to α, we find that there exists some element c_{n_0} such that $c_{n_0} > \alpha$ and such that the image $\psi(c_{n_0})$ lies to the left of P. Therefore, we have $c_{n_0} \in \mathfrak{M}$, that is, $c_{n_0} \leq \alpha$. But this is a contradiction, whence our assumption is false, and we have proved the surjectivity of ψ.

Exercise 5.3. Prove that $\psi(\alpha)$ is independent of the choice of sequence of nested intervals for $\alpha \in \mathbb{R}$.

We have now identified the set $\mathbb{R}$ of real numbers with the line G. We shall henceforth refer to this line as the *real number line*. A point P on the real number line determines an interval, namely the (directed) segment from the zero point P_0 to P. Addition and multiplication of real numbers translates, as in our discussion of rational numbers, to an addition and multiplication of the corresponding intervals. We can consider the real number line with these two operations to be a model for the field $\mathbb{R}$ of real numbers.

We close this section with a historical note. The axiom of geometric completeness that we required in the above discussion was characterized by Richard Dedekind in his 1872 essay *Continuity and Irrational Numbers*, in which he introduced the notion of what are now called *Dedekind cuts* with the following words:

> The above comparison of the domain $\mathcal{R}$ of rational numbers with a straight line has led to the recognition of the existence of gaps, of a certain incompleteness or discon-

tinuity of the former, while we ascribe to the straight line completeness, absence of gaps, or continuity. In what then does this continuity consist? Everything must depend on the answer to this question, and only through it shall we obtain a scientific basis for the investigation of *all* continuous domains. By vague remarks upon the unbroken connection in the smallest parts obviously nothing is gained; the problem is to indicate a precise characteristic of continuity that can serve as the basis for valid deductions. For a long time I pondered over this in vain, but finally I found what I was seeking. This discovery will, perhaps, be differently estimated by different people; the majority may find its substance very commonplace. It consists of the following. In the preceding section attention was called to the fact that every point p of the straight line produces a separation of the same into two portions such that every point of one portion lies to the left of every point of the other.[1]

6. The Axiomatic Point of View

Definition 6.1. A field $(K, +, \cdot)$ is said to be *ordered* if for all $\alpha \in K$, there exists a relation $\alpha > 0$ satisfying the following two properties:

(i) Exactly one of the following three possibilities holds: $\alpha > 0$, $\alpha = 0$, $\alpha < 0$ (that is, $-\alpha > 0$).

(ii) If we have $\alpha, \beta \in K$ with $\alpha, \beta > 0$, then we have $\alpha + \beta > 0$ and $\alpha \cdot \beta > 0$.

Remark 6.2. If $(K, +, \cdot)$ is an ordered field, then the order relation allows us to define an absolute value for the elements of K, which in turn allows us to introduce the notion of a Cauchy sequence $(\alpha_n) \subseteq K$.

Definition 6.3. An ordered field $(K, +, \cdot)$ is said to be *complete* if every Cauchy sequence $(\alpha_n) \subseteq K$ converges to an element of K.

Remark 6.4. The field $\mathbb{R}$ of real numbers is a complete ordered field. As in the case of the real numbers, completeness in an arbitrary ordered field can also be characterized in terms of the supremum and infimum principles or the nested intervals principle.

To complete the chapter, we sketch a proof of a theorem, which goes back to David Hilbert, stating that the field of real numbers $\mathbb{R}$ is (up to isomorphism) the unique complete ordered field.

Theorem 6.5. *A complete ordered field $(K, +, \cdot)$ is uniquely determined up to order-preserving ring isomorphism; that is, if $(K', +, \cdot)$ is an arbitrary complete ordered field, then there exists a ring isomorphism*

$$\varphi : (K, +, \cdot) \longrightarrow (K', +, \cdot)$$

[1] Translation by Wooster Woodruff Beman, 1901.

such that

$$\alpha > 0 \quad \Longrightarrow \quad \varphi(\alpha) > 0 \quad (\alpha \in K).$$

□

Remark 6.6. We state without proof that a complete ordered field K has a subring isomorphic to the ring $\mathbb{Z}$ of integers and therefore also a subfield isomorphic to the field $\mathbb{Q}$ of rational numbers. We shall henceforth identify $\mathbb{Z}$ and $\mathbb{Q}$ with this subring and subfield. We can then show that our complete ordered field K is *archimedean*; that is, it has the property that for every $\alpha, \beta \in K$ with $0 < \alpha < \beta$, there exists $n \in \mathbb{N}$ such that $n \cdot \alpha > \beta$. This leads to the fact that the rational numbers $\mathbb{Q}$ are dense in K; that is, there is a rational number r in every ε-neighborhood

$$U_\varepsilon = \{\beta \in K \mid \alpha - \varepsilon < \beta < \alpha + \varepsilon\}$$

of $\alpha \in K$.

Proof. We are now ready to begin our sketch of Theorem 6.5. We divide the proof into three steps.

Step 1: We have first to define a mapping $\varphi : K \longrightarrow K'$. To this end, we consider, for $\alpha \in K$, the set

$$\mathfrak{M}_\alpha := \{r \in \mathbb{Q} \mid r < \alpha\} \subseteq K.$$

Since $\mathbb{Q}$ is dense in K, the set $\mathfrak{M}_\alpha$ is nonempty. It is clearly bounded from above, and there exists, therefore, by the completeness of K, the supremum of $\mathfrak{M}_\alpha$, and we easily see that

$$\sup(\mathfrak{M}_\alpha) = \alpha.$$

By our earlier identification, we have the inclusions

$$\mathfrak{M}_\alpha \subseteq \mathbb{Q} \subseteq K'.$$

Since K' is also complete, there exists as well the supremum for the nonempty set $\mathfrak{M}_\alpha$ in K', which is bounded from above; we denote this supremum by $\sup'(\mathfrak{M}_\alpha)$. With this, we define the mapping $\varphi : K \longrightarrow K'$ by

$$\varphi(\alpha) := \sup'(\mathfrak{M}_\alpha).$$

One easily verifies that for $r \in \mathbb{Q}$, we have $\varphi(r) = r$. We can now see as follows that φ is order-preserving. If $\alpha, \beta \in K$ with $\alpha < \beta$, then there exists, because $\mathbb{Q}$ is dense in K, a number $r \in \mathbb{Q}$ such that $\alpha < r < \beta$; it follows that

$$\sup'(\mathfrak{M}_\alpha) < r < \sup'(\mathfrak{M}_\beta),$$

that is, $\varphi(\alpha) < \varphi(\beta)$, which proves that φ is order-preserving.

Step 2: We show here that the mapping φ is bijective. We begin with the proof of injectivity. If we have $\alpha, \beta \in K$ with $\alpha \neq \beta$, then we may assume without loss of generality that $\alpha < \beta$. Because φ is order-preserving, we have also $\varphi(\alpha) < \varphi(\beta)$, that is, $\varphi(\alpha) \neq \varphi(\beta)$, which proves that φ is injective.

To show the surjectivity of φ, we choose $\alpha' \in K'$ and consider the set $\mathfrak{M}_{\alpha'} \subseteq K'$, which we also, of course, view as a subset of K. We set $\alpha := \sup(\mathfrak{M}_{\alpha'}) \in K$. From the equality

$$\sup(\mathfrak{M}_\alpha) = \alpha = \sup(\mathfrak{M}_{\alpha'}),$$

we see that the sets $\mathfrak{M}_\alpha$ and $\mathfrak{M}_{\alpha'}$ are identical, whence

$$\varphi(\alpha) = \sup'(\mathfrak{M}_\alpha) = \sup'(\mathfrak{M}_{\alpha'}) = \alpha'.$$

This proves that φ is surjective.

Step 3: We now show that φ is a ring homomorphism. We begin with the additivity of φ. For $\alpha, \beta \in K$, we consider the set

$$\mathfrak{N}_{\alpha,\beta} := \{r + s \mid r, s \in \mathbb{Q}, r < \alpha, s < \beta\}$$

and show that $\mathfrak{N}_{\alpha,\beta} = \mathfrak{M}_{\alpha+\beta}$. Namely, if $t := r + s \in \mathfrak{N}_{\alpha,\beta}$, that is, $t = r + s$ with $r, s \in \mathbb{Q}$ and $r < \alpha$, $s < \beta$, then we have $t \in \mathbb{Q}$ and $t < \alpha + \beta$, that is, $t \in \mathfrak{M}_{\alpha+\beta}$. That is, we have the inclusion $\mathfrak{N}_{\alpha,\beta} \subseteq \mathfrak{M}_{\alpha+\beta}$. Conversely, if $t \in \mathfrak{M}_{\alpha+\beta}$, that is, $t \in \mathbb{Q}$ and $t < \alpha + \beta$, then using the density of $\mathbb{Q}$ in K, we can find $r \in \mathbb{Q}$ that satisfies

$$t - \beta < r < \alpha.$$

Setting $s := t - r$, we obtain rational numbers r, s with $r < \alpha$ and $s < \beta$. Since now $t = r + s$ with $r, s \in \mathbb{Q}$ and $r < \alpha$, $s < \beta$, we see that we have $t \in \mathfrak{N}_{\alpha,\beta}$, from which follows the inclusion $\mathfrak{M}_{\alpha+\beta} \subseteq \mathfrak{N}_{\alpha,\beta}$.

It follows from the equality of sets $\mathfrak{N}_{\alpha,\beta} = \mathfrak{M}_{\alpha+\beta}$ that

$$\varphi(\alpha + \beta) = \sup'(\mathfrak{M}_{\alpha+\beta}) = \sup'(\mathfrak{N}_{\alpha,\beta}). \tag{11}$$

It remains to show that $\sup'(\mathfrak{N}_{\alpha,\beta}) = \varphi(\alpha) + \varphi(\beta)$. To this end, we assert that $\varphi(\alpha) + \varphi(\beta)$ is an upper bound for $\mathfrak{N}_{\alpha,\beta}$. Indeed, if $t := r + s \in \mathfrak{N}_{\alpha,\beta}$, that is, $t = r + s$ with $r, s \in \mathbb{Q}$ and $r < \alpha$, $s < \beta$, then we have

$$t = r + s = \varphi(r) + \varphi(s) < \varphi(\alpha) + \varphi(\beta),$$

from which follows the assertion. It remains now to show that $\varphi(\alpha) + \varphi(\beta)$ is the least upper bound for $\mathfrak{N}_{\alpha,\beta}$. Let $K' \ni \gamma < \varphi(\alpha) + \varphi(\beta)$ be a smaller upper bound for $\mathfrak{N}_{\alpha,\beta}$. Since $\mathbb{Q}$ is dense in K', there exist $t, r \in \mathbb{Q}$ such that

$$\gamma < t < \varphi(\alpha) + \varphi(\beta), \quad t - \varphi(\beta) < r < \varphi(\alpha).$$

Setting $s := t - r$, we have obtained rational numbers r, s with $r < \alpha$ and $s < \beta$. Since now $t = r + s$ with $r, s \in \mathbb{Q}$ and $r < \alpha$, $s < \beta$, we see that $t \in \mathfrak{N}_{\alpha,\beta}$. But since we have also $\gamma < t$, it follows that γ cannot be an upper bound for $\mathfrak{N}_{\alpha,\beta}$. Therefore, we have

$$\varphi(\alpha) + \varphi(\beta) = \sup'(\mathfrak{N}_{\alpha,\beta}).$$

The additivity of φ now comes from (11). If $\alpha > 0$, then we obtain in particular from the fact that φ is additive and order-preserving that

$$\varphi(\alpha) > \varphi(0) = 0.$$

One proceeds similarly to prove the multiplicativity of φ. We shall omit the proof. This completes our sketch of the proof of Theorem 6.5. □

Exercise 6.7. Fill in the missing details in the above proof sketch.

Remark 6.8. If we proceed from the axiomatic point of view, it is not clear a priori that a complete ordered field K even exists. Our construction of the real numbers $\mathbb{R}$ in Section 2 has given us a model of such a field. An alternative model is provided by the real number line.

D. The p-adic Numbers: Another Completion of $\mathbb{Q}$

Following our construction of the real numbers, we end this chapter by introducing an alternative completion of the rational numbers that will lead us to the set of p-adic numbers. In what follows, we shall discuss the utility of the p-adic numbers, which will give us interesting new insights that will again lead us to some current problems in number theory.

D.1 The p-adic Absolute Value

We begin with the definition of the absolute value on the field $\mathbb{Q}$ of rational numbers.

Definition D.1. A mapping $\|\cdot\| : \mathbb{Q} \longrightarrow \mathbb{R}$ is called an *absolute value* if it satisfies the following three properties:
(i) For all $r \in \mathbb{Q}$, one has $\|r\| \geq 0$ and $\|r\| = 0 \iff r = 0$.
(ii) For all $r, s \in \mathbb{Q}$, one has the *product rule* $\|r \cdot s\| = \|r\| \cdot \|s\|$.
(iii) For all $r, s \in \mathbb{Q}$, one has the *triangle inequality* $\|r + s\| \leq \|r\| + \|s\|$.

Example D.2. (i) The absolute value $|\cdot|$ of a rational number introduced in Definition 6.8 of Chapter III is clearly an absolute value in the above sense. We call it the *archimedean absolute value on* $\mathbb{Q}$.

(ii) If we define $|r|_{\text{triv}} := 1$ for $r \in \mathbb{Q}$, $r \neq 0$, and set $|0|_{\text{triv}} := 0$, we obtain another absolute value $|\cdot|_{\text{triv}}$, which we call the *trivial absolute value on* $\mathbb{Q}$.

Remark D.3. We can construct yet another absolute value, one for each prime number p, in the following way: If $r = a/b$ is a nonzero rational number, we can write r in the form

$$r = \frac{a'}{b'}\, p^n,$$

where a', b' are nonzero integers relatively prime to p, and n is an integer (which by the fundamental theorem of arithmetic is uniquely determined). If we now set $v_p(r) := n$ and $v_p(0) := \infty$, we obtain a mapping

$$v_p : \mathbb{Q} \longrightarrow \mathbb{Z} \cup \{\infty\},$$

which clearly possesses the following two properties:
(i) For all $r, s \in \mathbb{Q}$, one has $v_p(r \cdot s) = v_p(r) + v_p(s)$.
(ii) For all $r, s \in \mathbb{Q}$, one has $v_p(r + s) \geq \min(v_p(r), v_p(s))$.
We call the mapping v_p the *p-adic valuation on* $\mathbb{Q}$. With it, we define for $r \in \mathbb{Q}$ the quantity

$$|r|_p := p^{-v_p(r)}.$$

Using the properties of the p-adic valuation $v_p(\cdot)$, one can verify at once that $|\cdot|_p$ defines an absolute value. In particular, the validity of the triangle inequality can be seen as follows:

$$|r + s|_p = p^{-v_p(r+s)} \leq p^{-\min(v_p(r), v_p(s))} = \max(|r|_p, |s|_p) \leq |r|_p + |s|_p,$$

that is, one has in fact the sharper inequality $|r + s|_p \leq \max(|r|_p, |s|_p)$, which is known as the *ultrametric inequality*.

Definition D.4. We call the absolute value $|\cdot|_p$ constructed in the previous remark for an arbitrary prime p the *p-adic absolute value on* $\mathbb{Q}$.

Remark D.5. If r is a nonzero rational number, then using the definition of the p-adic absolute value, one may easily verify the important formula

$$|r| \cdot \prod_{p \in \mathbb{P}} |r|_p = 1,$$

known as the *product formula*.

The larger the power n of the prime number p that divides $r \in \mathbb{Q}$, the smaller the p-adic absolute value $|r|_p$ of r, and conversely. This is illustrated in the following examples.

Example D.6. (i) Let

$$r = \frac{96}{9\,801} = \frac{2^5 \cdot 3^1}{3^4 \cdot 11^2}.$$

Then $v_2(r) = 5$, $v_3(r) = -3$, $v_{11}(r) = -2$, and $v_p(r) = 0$ for all primes p such that $p \neq 2, 3, 11$. It follows that $|r|_2 = \frac{1}{32}$, $|r|_3 = 27$, $|r|_{11} = 121$, and $|r|_p = 1$ for all primes p such that $p \neq 2, 3, 11$.

(ii) Let $r_1 = 735 = 3 \cdot 5 \cdot 7^2$, $r_2 = 3 \cdot 5 \cdot 7^{12}$, and $r_3 = 3 \cdot 5 \cdot 7^{-10}$. Then

$$|r_1|_7 = \frac{1}{49}, \quad |r_2|_7 = \frac{1}{7^{12}} = \frac{1}{13\,841\,287\,201}, \quad |r_3|_7 = 7^{10} = 282\,475\,249,$$

but $|r_1|_p = |r_2|_p = |r_3|_p$ for all primes p such that $p \neq 7$.

We recall that two absolute values $\|\cdot\|$ and $\|\cdot\|'$ are said to be equivalent if there exists a positive real number σ such that $\|\cdot\|' = \|\cdot\|^\sigma$. With this in mind, we obtain the following theorem, due to Alexander Ostrowski.

Theorem D.7 (Ostrowski [5]). *Every nontrivial absolute value*

$$\|\cdot\| : \mathbb{Q} \longrightarrow \mathbb{R}$$

defined on the field $\mathbb{Q}$ *of rational numbers is equivalent either to the archimedean absolute value* $|\cdot|$ *or to a p-adic absolute value* $|\cdot|_p$. □

Remark D.8. For two distinct primes p, q, the associated absolute values $|\cdot|_p$ and $|\cdot|_q$ are inequivalent.

D.2 The p-adic Numbers

In analogy to the notions of a rational Cauchy sequence and rational null sequence that we introduced in Definition 2.1 in relation to the archimedean absolute value, we can now also define rational Cauchy sequences and rational null sequences with respect to the p-adic absolute value.

Definition D.9. A sequence $(a_n) = (a_n)_{n \geq 0}$ with $a_n \in \mathbb{Q}$ for all $n \in \mathbb{N}$ is called a *rational Cauchy sequence with respect to the p-adic absolute value* if for every $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, there exists $N(\epsilon) \in \mathbb{N}$ such that for all $m, n \in \mathbb{N}$ with $m, n > N(\epsilon)$, the inequality

$$|a_m - a_n|_p < \epsilon$$

holds.

A sequence $(a_n) = (a_n)_{n \geq 0}$ with $a_n \in \mathbb{Q}$ for all $n \in \mathbb{N}$ is called a *rational null sequence with respect to the p-adic absolute value* if for every $\epsilon \in \mathbb{Q}$, $\epsilon > 0$,

there exists $N(\epsilon) \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n > N(\epsilon)$, the inequality

$$|a_n|_p < \epsilon$$

holds.

Example D.10. (i) The sequence $(a_n) = (7^n)_{n\geq 0}$ is a rational null sequence with respect to the 7-adic absolute value; to see this, one has only to observe that $|7^n|_7 = 7^{-n}$ holds for all $n \in \mathbb{N}$. If, on the other hand, p is a prime different from 7, then $(7^n)_{n\geq 0}$ is not a rational null sequence with respect to the p-adic absolute value, although it is a bounded sequence, since in this case, one has $|7^n|_p = 1$ for all $n \in \mathbb{N}$.

(ii) Let p be an arbitrary prime. Then the sequence $(a_n) = (\frac{1}{n})_{n>0}$ is not a rational null sequence with respect to the p-adic absolute value, since for the subsequence $(p^{-m})_{m\geq 0}$, one has

$$\Big|\frac{1}{p^m}\Big|_p = p^m.$$

This shows that the sequence $(a_n) = (\frac{1}{n})_{n>0}$ is not convergent.

(iii) The sequence $(a_n) = (2^{-n})_{n\geq 0}$ is a bounded sequence with respect to the 7-adic absolute value. It is not, however, a Cauchy sequence, since

$$\Big|\frac{1}{2^n} - \frac{1}{2^{n+1}}\Big|_7 = \Big|\frac{1}{2^{n+1}}\Big|_7 = 1.$$

These examples provide a first impression of a "calculus" of the p-adic absolute value, known as *p-adic analysis*, which appears to contradict much of our experience with real analysis. We refer the interested reader to the literature on the subject, in particular, the books [1] and [3].

Remark D.11. In analogy to the construction of the real numbers, let us now consider the set M_p of all rational Cauchy sequences with respect to the p-acid absolute value, that is,

$$M_p = \big\{(a_n) \,\big|\, (a_n) \text{ is a rational Cauchy sequence with respect to the } p\text{-adic absolute value}\big\}.$$

If we equip M_p with a componentwise additive operation $+$ and a multiplicative operation $\cdot$, then $(M_p, +, \cdot)$ becomes a commutative ring with unit element. Analogously, we set

$$\mathfrak{n}_p = \big\{(a_n) \in M_p \,\big|\, (a_n) \text{ is a rational null sequence with respect to the } p\text{-adic absolute value}\big\},$$

and we see that $\mathfrak{n}_p$ is an ideal of M_p. As in the case of the archimedean absolute value, one can show that the quotient ring $(M_p/\mathfrak{n}_p, +, \cdot)$ is a field.

Definition D.12. We call the field $(M_p/\mathfrak{n}_p, +, \cdot)$ the *field of p-adic numbers* and denote it by $\mathbb{Q}_p$.

The field of p-adic numbers was discovered at the end of the nineteenth century by Kurt Hensel.

Theorem D.13. *The mapping that associates with every rational number r the p-adic number $(r) + \mathfrak{n}_p$, where (r) denotes the rational Cauchy sequence with respect to the p-adic absolute value whose every term is equal to r, induces an injective ring homomorphism*

$$F_p : (\mathbb{Q}, +, \cdot) \longrightarrow (\mathbb{Q}_p, +, \cdot).$$

Moreover, the field $\mathbb{Q}_p$ of p-adic numbers is complete; that is, every Cauchy sequence $(\alpha_n) \subset \mathbb{Q}_p$ with respect to the p-adic absolute value converges to a limiting value $\alpha \in \mathbb{Q}_p$. □

Remark D.14. In analogy to the decimal fraction representation of a real number, it can be shown that a p-adic number $\alpha \in \mathbb{Q}_p$ can be represented by a series

$$\alpha = \sum_{j=\ell}^{\infty} q_j \, p^j$$

with $\ell \in \mathbb{Z}$ and $q_j \in \{0, \ldots, p-1\} = \mathbb{F}_p$. The set

$$\left\{ \alpha \in \mathbb{Q}_p \,\middle|\, \alpha = \sum_{j=0}^{\infty} q_j \, p^j \right\}$$

is called the set of *p-adic integers* and denoted by $\mathbb{Z}_p$; it can then be seen that $(\mathbb{Z}_p, +, \cdot)$ is a commutative subring of the field of p-adic numbers $(\mathbb{Q}_p, +, \cdot)$. From the definition of $\mathbb{Z}_p$, it can be seen at once that one has an isomorphism

$$\mathbb{Z}_p / p\mathbb{Z}_p \cong \mathbb{F}_p.$$

D.3 The Local–Global Principle

For a polynomial $P(X_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]$ in the n variables $X_1, \ldots, X_n$ with integer coefficients, we called in Appendix C an n-tuple of rational numbers $(x_1, \ldots, x_n)$ such that $P(x_1, \ldots, x_n) = 0$ a rational zero of the polynomial $P(X_1, \ldots, X_n)$. We had previously not investigated the question of

the existence of rational zeros of polynomials. Rather, we considered the question of the existence of finitely or infinitely many zeros of polynomials under the assumption that at least one such zero existed. In this section, we shall investigate the question of existence by considering some examples.

Remark D.15. As earlier, let $P(X_1,\dots,X_n) \in \mathbb{Z}[X_1,\dots,X_n]$, and let $(x_1,\dots,x_n) \in \mathbb{Q}^n$ be a rational zero of $P(X_1,\dots,X_n)$. From the embedding of $\mathbb{Q}$ in the p-adic numbers $\mathbb{Q}_p$ as well as in the real numbers $\mathbb{R}$, the n-tuple $(x_1,\dots,x_n)$ can also be viewed as an element of $\mathbb{Q}_p^n$ or of $\mathbb{R}^n$. We thereby observe that the existence of a rational zero of $P(X_1,\dots,X_n)$ entails the existence of a p-adic zero for all prime numbers p as well as a real zero. This, of course, raises at once the converse question: If a polynomial $P(X_1,\dots,X_n)$ has for all primes p a p-adic zero as well as a real zero, does the polynomial $P(X_1,\dots,X_n)$ then necessarily possess a rational zero?

To simplify the notation, we shall denote the set of real numbers $\mathbb{R}$ by $\mathbb{Q}_\infty$ and the archimedean absolute value $|\cdot|$ by $|\cdot|_\infty$.

Definition D.16. A polynomial $P(X_1,\dots,X_n) \in \mathbb{Z}[X_1,\dots,X_n]$ *satisfies the local–global principle* if the existence of p-adic zeros for all $p \in \mathbb{P} \cup \{\infty\}$ implies the existence of a rational zero of the polynomial $P(X_1,\dots,X_n)$.

If the polynomial $P(X_1,\dots,X_n)$ satisfies the local–global principle, then in the search for rational zeros, we are led to the search for p-adic zeros of the polynomial $P(X_1,\dots,X_n)$ for all $p \in \mathbb{P} \cup \{\infty\}$. We shall now consider this question for $p \in \mathbb{P}$. We begin by recalling the following simple lemma.

Lemma D.17. *Let $P(X_1,\dots,X_n) \in \mathbb{Z}[X_1,\dots,X_n]$ be a polynomial and $p \in \mathbb{P}$ a prime number. Then the following three statements are equivalent:*
(i) *$P(X_1,\dots,X_n)$ has a zero in $\mathbb{Q}_p^n$,*
(ii) *$P(X_1,\dots,X_n)$ has a primitive zero in $\mathbb{Z}_p^n$,*
(iii) *$P(X_1,\dots,X_n)$ has a zero in $(\mathbb{Z}/p^m\mathbb{Z})^n$ for all $m \in \mathbb{N}_{>0}$,*
where a zero $(x_1,\dots,x_n) \in \mathbb{Z}_p^n$ is primitive if not all the x_j $(j = 1,\dots,n)$ are divisible by p. □

This lemma reduces the question of p-adic zeros of a polynomial to the solution of polynomial congruences modulo p^m for all $m \in \mathbb{N}_{>0}$ (we refer in this regard to Section B.2). The following theorem shows under what conditions a zero modulo p already gives rise to a p-adic zero.

Theorem D.18. *Let $P(X_1,\dots,X_n) \in \mathbb{Z}[X_1,\dots,X_n]$ be a polynomial and $p \in \mathbb{P}$ a prime. Then every simple zero of $P(X_1,\dots,X_n)$ modulo p induces a p-adic zero of $P(X_1,\dots,X_n)$. Here a zero $(x_1,\dots,x_n) \in \mathbb{Z}^n$ modulo p is called simple if*

$$\left(\frac{\partial P}{\partial X_j}(x_1,\dots,x_n)\right)_{j=1,\dots,n} \not\equiv (0,\dots,0) \mod p.$$

Proof. We carry out the proof first for the case $n = 1$ and for simplicity of notation write $X = X_1$. Let $x^{(0)} \in \mathbb{Z}$ be a simple zero of the polynomial $P(X)$ modulo p. That is, we have

$$P(x^{(0)}) \equiv 0 \mod p, \text{ i.e., } P(x^{(0)}) = p\,a \text{ with } a \in \mathbb{Z},$$
$$P'(x^{(0)}) \not\equiv 0 \mod p, \text{ i.e., } P'(x^{(0)}) = b \text{ with } b \in \mathbb{Z}, (b,p) = 1.$$

With an integer y yet to be determined, we set $x^{(1)} := x^{(0)} + p\,y$ and obtain, with the help of Taylor's formula,

$$P(x^{(1)}) = P\big(x^{(0)}\big) + p\,y\,P'\big(x^{(0)}\big) + p^2\,c = p(a + y\,b) + p^2\,c,$$

for some $c \in \mathbb{Z}$. Since b is relatively prime to p, there exists $y \in \mathbb{Z}$ such that $a + y\,b \equiv 0 \mod p$. We thereby obtain

$$P(x^{(1)}) \equiv 0 \mod p^2 \quad \text{and} \quad x^{(1)} \equiv x^{(0)} \mod p.$$

If we proceed in the same way, that is, define $x^{(2)} := x^{(1)} + p^2 z$ with some $z \in \mathbb{Z}$ yet to be determined, we obtain a sequence of integers $(x^{(0)}, x^{(1)}, x^{(2)}, \dots)$ for which we have

$$P(x^{(j)}) \equiv 0 \mod p^{j+1} \quad \text{and} \quad x^{(j+1)} \equiv x^{(j)} \mod p^{j+1} \quad (j = 0,1,2,\dots).$$

The sequence $(x^{(0)}, x^{(1)}, x^{(2)}, \dots)$ is clearly a Cauchy sequence with respect to the p-adic absolute value that converges to a p-adic number, say ξ. By construction, we have $P(\xi) = 0$; that is, ξ is the desired p-adic zero.

Finally, we reduce the case $n > 1$ to the case $n = 1$. We begin with the simple zero $(x_1, \dots, x_n)$ of $P(X_1, \dots, X_n)$ modulo p and define for $k \in \{1, \dots, n\}$ the polynomial

$$Q(X_k) := P(x_1, \dots, x_{k-1}, X_k, x_{k+1}, \dots, x_n) \in \mathbb{Z}[X_k].$$

We then construct as previously, beginning with x_k, a p-adic zero ξ_k of $Q(X_k)$ (with a suitable choice of k). After interpreting the x_j as p-adic numbers ξ_j $(j = 1, \dots, n; j \neq k)$, we obtain, as desired, $P(\xi_1, \dots, \xi_n) = 0$. This completes the proof of the theorem. □

Remark D.19. The method employed in the previous proof of constructing a p-adic zero ξ of the polynomial $P(X)$ by beginning with a zero $x^{(0)} \in \mathbb{Z}$ modulo p yields in the case of the field $\mathbb{R}$ of real numbers the well-known Newton's method of finding the zeros of a polynomial.

D.4 The Theorem of Hasse–Minkowski

In this section, we shall introduce a special class of polynomials for which the local–global principle is valid. To this end, we consider the set

$$\mathcal{Q} := \left\{ Q(X_1, \ldots, X_n) = \sum_{j,k=1}^{n} a_{j,k} X_j X_k \,\middle|\, a_{j,k} = a_{k,j} \in \mathbb{Z}, \det(a_{j,k}) \neq 0 \right\}$$

of *nondegenerate quadratic forms over* $\mathbb{Z}$. We can now formulate the main theorem of this section.

Theorem D.20 (Hasse–Minkowski). *A quadratic form* $Q(X_1, \ldots, X_n) \in \mathcal{Q}$ *satisfies the local–global principle.* □

We shall not prove Theorem D.20 here; we refer the interested reader to the book [7]. Instead, we shall present some examples to indicate how one can use the Hasse–Minkowski theorem to find nontrivial rational zeros of nondegenerate quadratic forms over $\mathbb{Z}$.

Remark D.21. Ernst S. Selmer showed that the Hasse–Minkowski theorem cannot, in general, be generalized to cubic polynomials. To that end, he introduced the polynomial

$$P(X_1, X_2, X_3) = 3X_1^3 + 4X_2^3 + 5X_3^3,$$

which has nontrivial p-adic zeros for all $p \in \mathbb{P} \cup \{\infty\}$ but *no* nontrivial rational zeros. In contrast, Roger Heath-Brown showed that every cubic form in at least 14 variables satisfies the local–global principle (see [2]). According to an idea introduced by Yuri I. Manin, the obstruction to the validity of the local–global principle for cubic forms is measured by the Brauer group, a certain second cohomology group (see [4]). In general, further obstructions to the validity of the local–global principle for polynomials can arise. One speaks in this regard of the *Brauer–Manin Obstruction* (see [8]).

Remark D.22. By the Hasse–Minkowski theorem, a quadratic form $Q(X_1, \ldots, X_n) \in \mathcal{Q}$ has a nontrivial rational zero if the underlying form has nontrivial p-adic zeros for all $p \in \mathbb{P} \cup \{\infty\}$. This raises the question as to how one might prove the existence of such zeros. In the real case, that is, if $p = \infty$, this means simply that the given quadratic form must be indefinite. For the case $p \in \mathbb{P}$ with $p \neq 2$, we have the following sufficient criterion.

Proposition D.23. *A nondegenerate quadratic form*

$$Q(X_1, \ldots, X_n) = \sum_{j,k=1}^{n} a_{j,k} X_j X_k \in \mathcal{Q}$$

form has a nontrivial p-adic zero for all odd primes p if $n \geq 3$ and $v_p(\det(a_{j,k})) = 0$.

Proof. Since $v_p(\det(a_{j,k})) = 0$ holds for the odd prime p by hypothesis, the matrix $(a_{j,k})$ is invertible modulo p. Therefore, for all nontrivial integer n-tuples $(x_1,\ldots,x_n)$, one has the relationship

$$\left(\frac{\partial Q}{\partial X_j}(x_1,\ldots,x_n)\right)_{j=1,\ldots,n} = \left(\sum_{k=1}^{n} 2a_{j,k}x_k\right)_{j=1,\ldots,n} \not\equiv (0,\ldots,0) \mod p.$$

Thus the conditions of Theorem D.18 are satisfied, so that a nontrivial zero $(x_1,\ldots,x_n)$ modulo p can be lifted to a p-adic zero $(\xi_1,\ldots,\xi_n)$ of $Q(X_1,\ldots,X_n)$. We are therefore led to search for a nontrivial zero $(x_1,\ldots,x_n)$ modulo p of $Q(X_1,\ldots,X_n)$.

To that end, let us begin with the case $n = 3$. Then the quadratic form under consideration has, without loss of generality, the form

$$Q(X_1,X_2,X_3) = a_{1,1}X_1^2 + a_{2,2}X_2^2 + a_{3,3}X_3^2, \tag{12}$$

with $a_{j,j} \in \mathbb{Z}$ and $p \nmid a_{j,j}$ $(j = 1,2,3)$. We consider now the two sets

$$\mathcal{S}_1 := \{\overline{a}_{1,1}\overline{x}_1^2 \,|\, \overline{x}_1 \in \mathbb{F}_p\} \subseteq \mathbb{F}_p\,, \quad \mathcal{S}_2 := \{-\overline{a}_{2,2}\overline{x}_2^2 - \overline{a}_{3,3} \,|\, \overline{x}_2 \in \mathbb{F}_p\} \subseteq \mathbb{F}_p\,.$$

The sets $\mathcal{S}_1$ and $\mathcal{S}_2$ clearly have cardinality $(p+1)/2$, from which it follows that $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \varnothing$. There exists, therefore, a nontrivial zero of (12) of the form $(x_1,x_2,1)$ modulo p.

Finally, the case $n \geq 3$ can be easily reduced to the case $n = 3$, which completes the proof of the proposition. □

Remark D.24. The Hasse–Minkowski theorem shows, with the help of Proposition D.23, for example, that an indefinite unimodular (that is, $\det(a_{j,k}) = \pm 1$) form over $\mathbb{Z}$ of rank $n \geq 3$ has a nontrivial rational zero.

Having considered these examples of the existence of nontrivial rational zeros of nondegenerate quadratic forms over $\mathbb{Z}$, we shall end this section by introducing a necessary and sufficient criterion for the existence of nontrivial p-adic zeros of the quadratic form $Q(X_1,\ldots,X_n)$, which with the help of the Hasse–Minkowski theorem completely answers the question of the existence of nontrivial rational zeros of nondegenerate quadratic forms over $\mathbb{Z}$. But first, we recall the theory of quadratic residues and the law of quadratic reciprocity.

Definition D.25. Let p be a prime number and a an integer relatively prime to p. We define the *Legendre symbol* $\left(\frac{a}{p}\right)$ *of a over p* by

$$\left(\frac{a}{p}\right) := \begin{cases} +1, & \text{if } a \text{ is a quadratic residue modulo } p, \\ -1, & \text{if } a \text{ is a quadratic nonresidue modulo } p. \end{cases}$$

Here *a is a quadratic residue modulo p* means that there exists $x \in \mathbb{Z}$ such that $x^2 \equiv a \mod p$; otherwise, *a is a quadratic nonresidue modulo p*. If $p \mid a$, we set $\left(\frac{a}{p}\right) := 0$.

Remark D.26. Since the Legendre symbol is multiplicative with respect to the "numerator," its evaluation can be reduced to the cases $a = -1$, $a = 2$, and $a = q$ (q an odd prime). In the first two cases, one has

$$\left(\frac{-1}{p}\right) = \begin{cases} +1, & \text{if } p \equiv +1 \mod 4, \\ -1, & \text{if } p \equiv -1 \mod 4, \end{cases}$$

as well as

$$\left(\frac{2}{p}\right) = \begin{cases} +1, & \text{if } p \equiv \pm 1 \mod 8, \\ -1, & \text{if } p \equiv \pm 3 \mod 8. \end{cases}$$

The calculation of $\left(\frac{q}{p}\right)$ proceeds with the help of the law of quadratic reciprocity:

$$\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}} \left(\frac{p}{q}\right).$$

The Legendre symbol enables us to calculate the Hilbert symbol, which we now define for the field of p-adic numbers $\mathbb{Q}_p$.

Definition D.27. For $\alpha, \beta \in \mathbb{Q}_p$, we consider the quadratic form

$$-\alpha X_1^2 - \beta X_2^2 + X_3^2 \tag{13}$$

and define the *Hilbert symbol* $(\alpha, \beta)_p$ *of* α, β *with respect to* $\mathbb{Q}_p$ as $+1$ if (13) has a nontrivial solution $(x_1, x_2, x_3) \in \mathbb{Q}_p^3$ and as -1 otherwise.

Remark D.28. The Hilbert symbol $(\alpha, \beta)_p$ can be calculated as follows. We write

$$\begin{aligned} \alpha &= u \cdot p^a, \quad \text{with } u \in \mathbb{Z}_p^\times \text{ and } a \in \mathbb{Z}, \\ \beta &= v \cdot p^b, \quad \text{with } v \in \mathbb{Z}_p^\times \text{ and } b \in \mathbb{Z}; \end{aligned}$$

here we can identify u and v with integers relatively prime to p on account of the isomorphism $\mathbb{Z}_p / p\mathbb{Z}_p \cong \mathbb{Z}/p\mathbb{Z}$; we again denote these integers by u and v. Then if p is odd, we have

$$(\alpha,\beta)_p = \left(\frac{-1}{p}\right)^{ab}\left(\frac{u}{p}\right)^{b}\left(\frac{v}{p}\right)^{a}.$$

In the case $p = 2$, there is an equally simple formula.

Definition D.29. A nondegenerate quadratic form

$$Q(X_1,\dots,X_n) = \sum_{j,k=1}^{n} a_{j,k}X_jX_k$$

defined over $\mathbb{Z}$ has two important invariants in addition to its *degree* n.

The first invariant is given by the *discriminant* $\operatorname{disc}(Q)$ *of* Q, defined by $\det(a_{j,k}) \mod (\mathbb{Q}^\times)^2$.

The second invariant is given by the following collection of Hilbert symbols $\operatorname{hilb}_p(Q) \in \{\pm 1\}$ for $p \in \mathbb{P} \cup \{\infty\}$: if we diagonalize the quadratic form $Q(X_1,\dots,X_n)$ by a suitable choice of basis, we may assume without loss of generality that

$$Q(X_1,\dots,X_n) = a_1X_1^2 + \cdots + a_nX_n^2.$$

We thereby set

$$\operatorname{hilb}_p(Q) := \prod_{j<k} (a_j,a_k)_p.$$

It can be shown that $\operatorname{hilb}_p(Q)$ is equal to -1 for only a finite (even) number of primes (including $p = \infty$) and that it satisfies the relation

$$\prod_{p\in\mathbb{P}\cup\{\infty\}} \operatorname{hilb}_p(Q) = +1.$$

We can now formulate the promised necessary and sufficient criterion.

Theorem D.30. *A nondegenerate quadratic form*

$$Q(X_1,\dots,X_n) = a_1X_1^2 + \cdots + a_nX_n^2$$

defined over $\mathbb{Q}_p$ *has a nontrivial p-adic zero if and only if one of the following conditions holds:*

(i) $n = 2$ *and* $\operatorname{disc}(Q) = -1$.
(ii) $n = 3$ *and* $\operatorname{hilb}_p(Q) = (-1,-\operatorname{disc}(Q))_p$.
(iii) $n = 4$ *and* $\operatorname{disc}(Q) \neq +1$ *or* $\operatorname{disc}(Q) = +1$ *and* $\operatorname{hilb}_p(Q) = (-1,-1)_p$.
(iv) $n \geq 5$.

Proof. For the proof, we refer the reader to the book [7]. □

Remark D.31. We note that Theorem D.30 confirms the result given in Proposition D.23, since all Hilbert symbols appearing there are equal to $+1$.

This brings to an end this first look into the theory of p-adic numbers with examples that attest to its importance. In addition to the examples we have given, there are numerous further questions—some topics of current research—that can be initially investigated with p-adic methods, in the hope that the p-adic information obtained can lead to a global solution of the given problem.

References

[1] F. Q. Gouvêa: *p-adic numbers: an introduction.* Springer, Berlin Heidelberg New York, 2nd edition, 1997.

[2] D. R. Heath-Brown: *Cubic forms in* 14 *variables.* Invent. Math. **170** (2007), 199–230.

[3] N. Koblitz: *p-adic numbers, p-adic analysis, and zeta functions.* Springer, Berlin Heidelberg New York, 2nd edition, 1984.

[4] Y. I. Manin: *Cubic forms.* North-Holland Mathematical Library, Volume 4. North-Holland, Amsterdam, 2nd edition, 1986.

[5] A. Ostrowski: *Über einige Lösungen der Funktionalgleichung* $\varphi(x)\varphi(y) = \varphi(xy)$. Acta Math. **41** (1918), 271–284.

[6] E. S. Selmer: *The diophantine equation* $ax^3 + by^3 + cz^3 = 0$. Acta Math. **85** (1957), 203–362.

[7] J.-P. Serre: *A course in arithmetic.* Translated from the French original. Springer, Berlin Heidelberg New York, 1973.

[8] A. Skorobogatov: *Torsors and rational points.* Cambridge Tracts in Mathematics 144. Cambridge University Press, Cambridge, 2001.

V The Complex Numbers

1. The Complex Numbers as a Real Vector Space

Through the extension of the set of natural numbers to the integers and then to the field of rational numbers, it became possible to solve the linear equation

$$a \cdot x + b = c \qquad (a, b, c \in \mathbb{Q},\ a \neq 0).$$

A natural question that arises is whether and how we might be able to solve quadratic equations, in particular the purely quadratic (i.e., no linear term) equation

$$x^2 = a \tag{1}$$

for $a \in \mathbb{Q}$. If $a < 0$, then the equation is a priori unsolvable for $x \in \mathbb{Q}$, since the square of a rational number is always nonnegative. Moreover, even for $a > 0$, the equation is not necessarily solvable in rational numbers, as the example $a = 2$ shows. If there were such a solution, then there would exist positive natural numbers m, n such that

$$\frac{m^2}{n^2} = 2 \quad \Longleftrightarrow \quad m^2 = 2 \cdot n^2.$$

If we now take the prime factorizations of m and n, we see that on the left-hand side, all the prime factors appear to even powers, while the prime number 2 on the right-hand side appears to an odd power, which contradicts unique factorization.

With our extension of the rational numbers $\mathbb{Q}$ to $\mathbb{R}$, equation (1) is solvable for $a > 0$. Indeed, such an equation can be solved for any positive real number α, for if we have $\alpha \in \mathbb{R}$, $\alpha > 0$, then as we shall now see, the purely quadratic equation

$$x^2 = \alpha$$

has a real solution. We begin by choosing a real positive number β_0 and define recursively, for $n \in \mathbb{N}$,

$$\beta_{n+1} := \frac{\alpha + \beta_n^2}{2\beta_n}. \tag{2}$$

One can easily verify that this defines a monotonically decreasing sequence (β_n) that is bounded from below. By the completeness of $\mathbb{R}$, this real sequence has a limit, namely

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0_5

$$\beta := \lim_{n\to\infty} \beta_n = \inf_{n\in\mathbb{N}} \{\beta_n\}.$$

By passing to the limit on both sides of (2), we see that $\beta^2 = \alpha$. We notate this by writing $\beta = \sqrt{\alpha}$ and note that there is an additional solution, namely $\beta = -\sqrt{\alpha}$.

On the other hand, the equation $x^2 = \alpha$ remains unsolvable for $\alpha < 0$ in the field of real numbers, as can be seen in the example $x^2 = -1$. We shall therefore now undertake to enlarge the field $\mathbb{R}$ of real numbers in such a way that the quadratic equation $x^2 = -1$ has a solution. And we shall do so in such a way that the extension that we create is again a field.

Exercise 1.1. Using the above procedure, compute $\sqrt{3}$ and $\sqrt{5}$ to ten decimal places.

Definition 1.2. We define $i := \sqrt{-1}$, that is, $i^2 = -1$. We call i the *imaginary unit*.

We now use the imaginary unit i to define the *complex numbers*.

Definition 1.3. The set of *complex numbers* $\mathbb{C}$ is defined to be the set of all real linear combinations of the unit element 1 in $\mathbb{R}$ and the imaginary unit i. That is, we have

$$\mathbb{C} := \{\alpha = \alpha_1 \cdot 1 + \alpha_2 \cdot i \mid \alpha_1, \alpha_2 \in \mathbb{R}\}.$$

In place of $\alpha = \alpha_1 \cdot 1 + \alpha_2 \cdot i \in \mathbb{C}$, we use the shorthand notation $\alpha_1 + \alpha_2 i$. The real number α_1 is called the *real part* of α and is denoted by $\mathrm{Re}(\alpha)$. The real number α_2 is called the *imaginary part* of α and is denoted by $\mathrm{Im}(\alpha)$. If $\mathrm{Re}(\alpha) = 0$, then we say that α is *purely imaginary*.

Remark 1.4. We can view the set $\mathbb{C}$ of complex numbers as a 2-dimensional real vector space with basis $\{1, i\}$. In this way, we can identify $\mathbb{C}$ with the Cartesian plane, which we call the *complex plane* in this case.

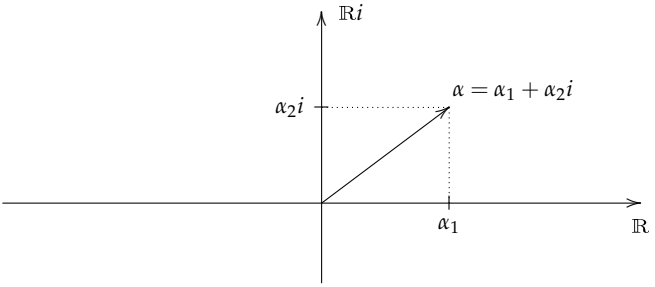


Fig. 1. The complex plane.

Remark 1.5. As an $\mathbb{R}$-vector space, $\mathbb{C}$ has in particular the structure of an abelian group. Namely, if $\alpha = \alpha_1 + \alpha_2 i$, $\beta = \beta_1 + \beta_2 i \in \mathbb{C}$, then we have

$$\alpha + \beta = (\alpha_1 + \alpha_2 i) + (\beta_1 + \beta_2 i) = (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2)i.$$

This addition is associative and commutative. The additive identity element, that is, the zero element of $\mathbb{C}$, is given by $0 := 0 + 0i$. If $\alpha = \alpha_1 + \alpha_2 i \in \mathbb{C}$, then the additive inverse of α is given by

$$-\alpha := (-\alpha_1) + (-\alpha_2)i = -\alpha_1 - \alpha_2 i.$$

Definition 1.6. The product of two complex numbers $\alpha = \alpha_1 + \alpha_2 i$ and $\beta = \beta_1 + \beta_2 i$ is given by

$$\alpha \cdot \beta = (\alpha_1 + \alpha_2 i) \cdot (\beta_1 + \beta_2 i) := (\alpha_1 \cdot \beta_1 - \alpha_2 \cdot \beta_2) + (\alpha_1 \cdot \beta_2 + \alpha_2 \cdot \beta_1)i.$$

Theorem 1.7. *The mathematical structure $(\mathbb{C}, +, \cdot)$ is a field with unit element $1 := 1 + 0i$ that contains the field $(\mathbb{R}, +, \cdot)$ of real numbers as a subfield.*

Furthermore, the quadratic equation

$$\alpha \cdot x^2 + \beta \cdot x + \gamma = 0, \tag{3}$$

with $\alpha, \beta, \gamma \in \mathbb{R}$, has solutions in $\mathbb{C}$.

Proof. We have already shown that $(\mathbb{C}, +)$ is an abelian group with additive identity element 0. It is also easy to verify that the multiplication of two complex numbers that we have defined is associative and commutative and that the unit element 1 is the multiplicative identity element. Once we have verified the two distributive laws

$$\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma, \quad (\beta + \gamma) \cdot \alpha = \beta \cdot \alpha + \gamma \cdot \alpha$$

for $\alpha, \beta, \gamma \in \mathbb{C}$, we will have shown that $(\mathbb{C}, +, \cdot)$ is a commutative ring with unit element 1. To prove the field property of $\mathbb{C}$, it remains to show that every $\alpha = \alpha_1 + \alpha_2 i \neq 0$ has a multiplicative inverse in $\mathbb{C}$. Since $\alpha \neq 0$, we must have $\alpha_1 \neq 0$ or $\alpha_2 \neq 0$, whence we have $\alpha_1^2 + \alpha_2^2 \neq 0$; using this fact, we can see easily that

$$\beta = \frac{\alpha_1}{\alpha_1^2 + \alpha_2^2} - \frac{\alpha_2}{\alpha_1^2 + \alpha_2^2} i$$

is the multiplicative inverse of α.

By the mapping

$$\psi : (\mathbb{R}, +, \cdot) \longrightarrow (\mathbb{C}, +, \cdot),$$

given by the assignment $\alpha_1 \mapsto \alpha_1 + 0i$, we obtain an injective ring homomorphism from $(\mathbb{R}, +, \cdot)$ to $(\mathbb{C}, +, \cdot)$. This shows that we may consider the field of real numbers to be a subfield of the field of complex numbers.

The quadratic equation (3) has the two solutions

$$x_{1,2} = \frac{-\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha},$$

where $\sqrt{\beta^2 - 4\alpha\gamma} = \sqrt{|\beta^2 - 4\alpha\gamma|}\, i$ if $\beta^2 - 4\alpha\gamma < 0$. This completes the proof of the theorem. □

Exercise 1.8. Fill in the missing details in the proof of Theorem 1.7.

Remark 1.9. One can generalize the previous theorem to show that the quadratic equation (3) with *complex* coefficients α, β, γ also has a solution in the field of complex numbers. This is an astounding result. We enlarged the field of real numbers to the complex numbers by introducing a single additional number, namely the square root of -1, and the result is that *every* quadratic equation with complex coefficients becomes solvable in $\mathbb{C}$.

Exercise 1.10. Derive a formula for the solutions of the quadratic equation $x^2 = \alpha$ for $\alpha = \alpha_1 + \alpha_2 i \in \mathbb{C}$, $\alpha \neq 0$. Use your formula to calculate the roots of $x^2 = \alpha$ for $\alpha = i$, $\alpha = 2 + i$, and $\alpha = 3 - 2i$.

Exercise 1.11. Find all solutions of the quadratic equations $x^2 + (1 + i) \cdot x + i = 0$ and $x^2 + (2 - i) \cdot x - 2i = 0$.

Definition 1.12. For $\alpha = \alpha_1 + \alpha_2 i \in \mathbb{C}$, we define the *complex conjugate* of α, denoted by $\overline{\alpha}$, as

$$\overline{\alpha} := \alpha_1 - \alpha_2 i.$$

In the complex plane, the point $\overline{\alpha}$ can be located by reflecting α in the real axis.

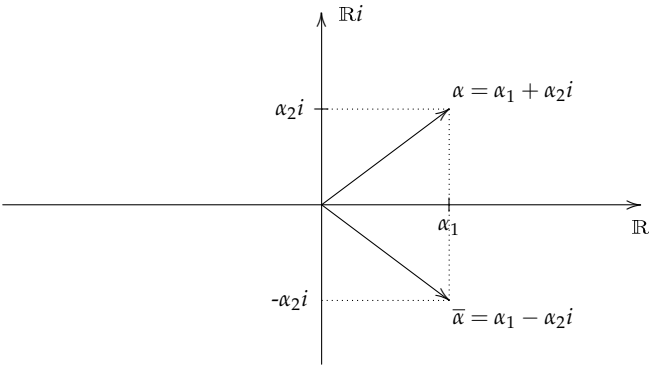


Fig. 2. The complex conjugate $\overline{\alpha}$ of the complex number α.

Definition 1.13. The Euclidean scalar product $\langle \cdot, \cdot \rangle : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{R}$ is defined by

$$\langle \alpha, \beta \rangle := \mathrm{Re}(\alpha \cdot \overline{\beta}) = \alpha_1\beta_1 + \alpha_2\beta_2,$$

where $\alpha = \alpha_1 + \alpha_2 i$, $\beta = \beta_1 + \beta_2 i \in \mathbb{C}$. The *absolute value*, or *modulus*, $|\alpha|$ of α is given by

$$|\alpha| := \sqrt{\alpha \cdot \overline{\alpha}} = \sqrt{\alpha_1^2 + \alpha_2^2}.$$

Exercise 1.14.
(a) Prove that for all $\alpha, \beta \in \mathbb{C}$, one has the *product rule* $|\alpha \cdot \beta| = |\alpha| \cdot |\beta|$.
(b) Use part (a) of this exercise to prove the following: if each of two natural numbers can be represented as the sum of the squares of two natural numbers, then the product of these two numbers can also be represented as the sum of the squares of two natural numbers.

Remark 1.15. Using the previous definition, the multiplicative inverse of $0 \neq \alpha \in \mathbb{C}$ can be written in the form

$$\alpha^{-1} = \frac{\overline{\alpha}}{|\alpha|^2}.$$

Moreover, one can easily check that the modulus function $|\cdot|$ has the properties of a norm. It turns out that the field $(\mathbb{C}, +, \cdot)$ is complete with respect to this norm.

2. Complex Numbers of Modulus 1 and the Special Orthogonal Group

In this section, we are going to identify the set of complex numbers of unit modulus with the special orthogonal group.

We begin by considering the noncommutative ring $(\mathrm{M}_2(\mathbb{R}), +, \cdot)$, which consists of all 2×2 matrices with real entries,

$$\mathrm{M}_2(\mathbb{R}) := \left\{ A = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix} \,\middle|\, \alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R} \right\},$$

together with matrix addition and multiplication as it has been defined in Example 2.4 (iv) of Chapter III. We denote the unit element of $\mathrm{M}_2(\mathbb{R})$ by

$$E := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

If $A = \left(\begin{smallmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{smallmatrix}\right) \in \mathrm{M}_2(\mathbb{R})$, then the *transpose* of A, denoted by A^t, is given by

$$A^t := \begin{pmatrix} \alpha_1 & \alpha_3 \\ \alpha_2 & \alpha_4 \end{pmatrix} \in \mathrm{M}_2(\mathbb{R}).$$

Definition 2.1. We define

$$\mathbb{S}^1 := \{\alpha \in \mathbb{C} \mid |\alpha| = 1\}$$

to be the set of all complex numbers of modulus 1.

Remark 2.2. The set $\mathbb{S}^1$ is a subgroup of the group $(\mathbb{C} \setminus \{0\}, \cdot)$.

Exercise 2.3. Verify the assertion of Remark 2.2.

In the following, we shall identify the field of complex numbers with a subring of the noncommutative ring $(M_2(\mathbb{R}), +, \cdot)$.

Lemma 2.4. *The mapping* $f : (\mathbb{C}, +, \cdot) \longrightarrow (M_2(\mathbb{R}), +, \cdot)$ *defined by*

$$\alpha = \alpha_1 + \alpha_2 i \mapsto \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix}$$

is an injective ring homomorphism. The image

$$\mathcal{C} := \operatorname{im}(f) = \left\{ \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix} \,\middle|\, \alpha_1, \alpha_2 \in \mathbb{R} \right\}$$

is a field. In particular, f *induces the isomorphism* $\mathbb{C} \cong \mathcal{C}$.

Proof. We begin by proving that f is an injective ring homomorphism. Let $\alpha = \alpha_1 + \alpha_2 i \in \mathbb{C}$ and $\beta = \beta_1 + \beta_2 i \in \mathbb{C}$. We obtain

$$\begin{aligned} f(\alpha + \beta) = f((\alpha_1 + \beta_1) + (\alpha_2 + \beta_2)i) &= \begin{pmatrix} \alpha_1 + \beta_1 & \alpha_2 + \beta_2 \\ -(\alpha_2 + \beta_2) & \alpha_1 + \beta_1 \end{pmatrix} \\ &= \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix} + \begin{pmatrix} \beta_1 & \beta_2 \\ -\beta_2 & \beta_1 \end{pmatrix} = f(\alpha) + f(\beta) \end{aligned}$$

and

$$\begin{aligned} f(\alpha \cdot \beta) &= f((\alpha_1 \cdot \beta_1 - \alpha_2 \cdot \beta_2) + (\alpha_1 \cdot \beta_2 + \alpha_2 \cdot \beta_1)i) \\ &= \begin{pmatrix} \alpha_1 \cdot \beta_1 - \alpha_2 \cdot \beta_2 & \alpha_1 \cdot \beta_2 + \alpha_2 \cdot \beta_1 \\ -(\alpha_1 \cdot \beta_2 + \alpha_2 \cdot \beta_1) & \alpha_1 \cdot \beta_1 - \alpha_2 \cdot \beta_2 \end{pmatrix} \\ &= \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 & \beta_2 \\ -\beta_2 & \beta_1 \end{pmatrix} = f(\alpha) \cdot f(\beta). \end{aligned}$$

Since we have $\ker(f) = \{0\}$, we have shown that f is an injective ring homomorphism. The image of f is the set

$$\mathcal{C} = \operatorname{im}(f) = \left\{ \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix} \,\middle|\, \alpha_1, \alpha_2 \in \mathbb{R} \right\}.$$

By Lemma 3.4 of Chapter III, $\mathcal{C}$ is in fact a subring of $(M_2(\mathbb{R}), +, \cdot)$. Finally, it follows from the homomorphism theorem for rings that we have the iso-

morphism

$$\mathbb{C} = \mathbb{C}/\ker(f) \cong \operatorname{im}(f) = \mathcal{C}.$$

But since $\mathbb{C}$ is a field, $\mathcal{C}$ must also be a field, as asserted. □

Exercise 2.5. Show that there are infinitely many subrings of $(\mathrm{M}_2(\mathbb{R}), +, \cdot)$ that are isomorphic to $\mathbb{C}$.

Definition 2.6. The *orthogonal group* $\mathrm{O}_2(\mathbb{R})$ is defined by

$$\mathrm{O}_2(\mathbb{R}) := \{A \in \mathrm{M}_2(\mathbb{R}) \mid A \cdot A^t = E\}.$$

The *special orthogonal group* $\mathrm{SO}_2(\mathbb{R})$ is defined by

$$\mathrm{SO}_2(\mathbb{R}) := \{A \in \mathrm{O}_2(\mathbb{R}) \mid \det(A) = 1\}.$$

Remark 2.7. The orthogonal group $(\mathrm{O}_2(\mathbb{R}), \cdot)$ is a group whose operation is matrix multiplication. The special orthogonal group $\mathrm{SO}_2(\mathbb{R})$ is a subgroup, indeed a normal subgroup, of $(\mathrm{O}_2(\mathbb{R}), \cdot)$.

Exercise 2.8. Show that we have $\det(A) = \pm 1$ for $A \in \mathrm{O}_2(\mathbb{R})$, and verify the assertions in Remark 2.7.

Theorem 2.9. *We have the group isomorphism*

$$(\mathbb{S}^1, \cdot) \cong (\mathrm{SO}_2(\mathbb{R}), \cdot).$$

Proof. We note first that for an arbitrary matrix

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix} \in \mathcal{C},$$

we have the equalities $\det(A) = \alpha_1^2 + \alpha_2^2$ and

$$A \cdot A^t = \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 & -\alpha_2 \\ \alpha_2 & \alpha_1 \end{pmatrix} = \begin{pmatrix} \alpha_1^2 + \alpha_2^2 & 0 \\ 0 & \alpha_1^2 + \alpha_2^2 \end{pmatrix} = \det(A) \cdot E.$$

If we now have $\alpha = \alpha_1 + \alpha_2 i \in \mathbb{S}^1$, that is, $|\alpha| = 1$, then it follows that $|\alpha|^2 = \alpha_1^2 + \alpha_2^2 = 1$, and we therefore obtain under the mapping f from Lemma 2.4, for

$$A = f(\alpha) = f(\alpha_1 + \alpha_2 i) = \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix},$$

the equalities $\det(A) = \alpha_1^2 + \alpha_2^2 = 1$ and $A \cdot A^t = \det(A) \cdot E = E$. This proves that $A \in \mathrm{SO}_2(\mathbb{R})$. Thus the mapping f induces an injective group homomorphism $g := f|_{\mathbb{S}^1} : \mathbb{S}^1 \longrightarrow \mathrm{SO}_2(\mathbb{R})$ with image

$$\operatorname{im}(g) = \left\{ A \in \mathcal{C} \,\middle|\, \det(A) = 1 \right\} \subseteq \mathrm{SO}_2(\mathbb{R}).$$

To prove the surjectivity of g, we have to prove that $\mathrm{SO}_2(\mathbb{R}) \subseteq \operatorname{im}(g)$. To this end, let

$$B := \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix} \in \mathrm{SO}_2(\mathbb{R}).$$

We then have $B \cdot B^t = E$, whence also $B^{-1} = B^t$. Since $\det(B) = 1$, we have also

$$B^{-1} = \begin{pmatrix} \alpha_4 & -\alpha_2 \\ -\alpha_3 & \alpha_1 \end{pmatrix}.$$

We must therefore have $\alpha_4 = \alpha_1$ and $\alpha_3 = -\alpha_2$, which proves that $B \in \mathcal{C}$. Since $\det(B) = 1$, it then follows that

$$\mathrm{SO}_2(\mathbb{R}) \subseteq \left\{ A \in \mathcal{C} \,\middle|\, \det(A) = 1 \right\} = \operatorname{im}(g).$$

This completes the proof. □

Corollary 2.10. *Every complex number $\alpha \in \mathbb{C} \setminus \{0\}$ can be represented uniquely in the form*

$$\alpha = |\alpha| \cdot (\cos(\varphi) + i \sin(\varphi)) \tag{4}$$

for some $\varphi \in [0, 2\pi)$.

Proof. The proof of Theorem 2.9 shows in particular that every matrix $A \in \mathrm{SO}_2(\mathbb{R})$ can be represented in the form

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix}$$

with $\alpha_1, \alpha_2 \in \mathbb{R}$, where we have the relation $\alpha_1^2 + \alpha_2^2 = 1$. We must also have $\alpha_1, \alpha_2 \in [-1, 1]$, and so there exists a uniquely determined $\varphi \in [0, 2\pi)$ with $\alpha_1 = \cos(\varphi)$ and $\alpha_2 = \sin(\varphi)$. Since we have $\alpha / |\alpha| \in \mathbb{S}^1$, there exists a uniquely determined $\varphi \in [0, 2\pi)$ with

$$\frac{\alpha}{|\alpha|} = \cos(\varphi) + i \sin(\varphi),$$

from which the assertion of the corollary follows. □

Remark 2.11. The representation (4) is called the *polar-coordinate* representation of the complex number α. Using this representation, one can prove the important fact that it is possible to extract the kth root of a complex number α for every $k \in \mathbb{N}$, $k > 1$. We saw this earlier, in Exercise 1.10, for the case $k = 2$.

Exercise 2.12. Prove *de Moivre's theorem*: Let $\alpha \in \mathbb{C} \setminus \{0\}$ be represented in polar coordinates as $\alpha = |\alpha| \cdot (\cos(\varphi) + i\sin(\varphi))$. Then we have the equality

$$\alpha^{m/n} = |\alpha|^{m/n} \cdot \Big(\cos\Big(\frac{m}{n}\varphi\Big) + i\sin\Big(\frac{m}{n}\varphi\Big)\Big)$$

for $m, n \in \mathbb{N}$ and $n \neq 0$.

3. The Fundamental Theorem of Algebra

In this section, we shall give an elementary proof of the fundamental theorem of algebra. In doing so, we shall use without proof the well-known fact from calculus that a continuous real function of one or several variables defined on a closed, bounded set assumes a minimum value.

Theorem 3.1 (Fundamental theorem of algebra). *Every polynomial*

$$f(X) = \alpha_n X^n + \alpha_{n-1} X^{n-1} + \cdots + \alpha_1 X + \alpha_0$$

of degree $n > 0$ with complex coefficients $\alpha_0, \ldots, \alpha_n$ has at least one root in the field $\mathbb{C}$. As a result, the polynomial f can be decomposed in $\mathbb{C}$ into linear factors. That is, there exist complex numbers $\zeta_1, \ldots, \zeta_n$ such that

$$f(X) = \alpha_n \cdot (X - \zeta_1) \cdots (X - \zeta_n).$$

Proof. Without loss of generality, we shall assume that $\alpha_n = 1$, and we write

$$f(X) = X^n + g(X),$$

with $g(X) := \alpha_{n-1} X^{n-1} + \cdots + \alpha_1 X + \alpha_0$. We begin by showing that there exists a complex number $\zeta_0 \in \mathbb{C}$ such that

$$|f(\zeta_0)| \leq |f(\zeta)|$$

for all $\zeta \in \mathbb{C}$. For the real number

$$r := 1 + |\alpha_{n-1}| + \cdots + |\alpha_1| + |\alpha_0| \in \mathbb{R},$$

we obtain for all $\zeta \in \mathbb{C}$ with $|\zeta| > r \geq 1$ the estimate

$$\begin{aligned} |g(\zeta)| &\leq |\zeta^{n-1}| \cdot \left(|\alpha_{n-1}| + \cdots + \frac{|\alpha_1|}{|\zeta^{n-2}|} + \frac{|\alpha_0|}{|\zeta^{n-1}|} \right) \\ &\leq |\zeta^{n-1}| \cdot \big(|\alpha_{n-1}| + \cdots + |\alpha_1| + |\alpha_0| \big) \\ &\leq |\zeta|^{n-1} \cdot (r-1) < |\zeta|^{n-1} \cdot (|\zeta| - 1). \end{aligned}$$

This yields for $\zeta \in \mathbb{C}$ with $|\zeta| > r \geq 1$ the inequality

$$|f(\zeta)| = |\zeta^n + g(\zeta)| \geq |\zeta^n| - |g(\zeta)|$$
$$\geq |\zeta|^n - |\zeta|^{n-1} \cdot (|\zeta| - 1) = |\zeta|^{n-1} \geq |\zeta| > r. \tag{5}$$

To obtain an estimate for $\zeta \in \mathbb{C}$ with $|\zeta| \leq r$, we identify $\mathbb{C}$ with the Cartesian plane, as in Remark 1.4. By decomposing the (complex) variable X and thereby also $f(X)$ into its real and imaginary parts, we may view f as a continuous real mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$. But then the function $|f|$ defined on the closed disk $\{\zeta \in \mathbb{C} \mid |\zeta| \leq r\} \subseteq \mathbb{R}^2$ must attain a minimum. That is, there exists $\zeta_0 \in \mathbb{C}$, $|\zeta_0| \leq r$, such that

$$|f(\zeta_0)| \leq |f(\zeta)| \tag{6}$$

for all $\zeta \in \mathbb{C}$ with $|\zeta| \leq r$. In particular, we must have

$$|f(\zeta_0)| \leq |f(0)| = |\alpha_0| < r. \tag{7}$$

In sum, the estimates (5), (6), and (7), prove that the inequality $|f(\zeta_0)| \leq |f(\zeta)|$ holds for all $\zeta \in \mathbb{C}$, as asserted.

We show now that ζ_0 is a zero of $f(X)$. Without loss of generality, we may assume that $\zeta_0 = 0$, since otherwise, we could as well consider the polynomial $f(X + \zeta_0)$. We shall carry out a proof by contradiction and assume that $f(0) = \alpha_0 \neq 0$. If $k \in \{1, \ldots, n\}$ is minimal with $\alpha_k \neq 0$, then we can write

$$f(X) = X^{k+1} \cdot h(X) + \alpha_k X^k + \alpha_0$$

for some polynomial $h(X)$. Since we may extract the kth root of any complex number, there exists a complex number $\beta \in \mathbb{C}$, $\beta \neq 0$, such that

$$\beta^k = -\frac{\alpha_0}{\alpha_k}.$$

We then define, for $t \in \mathbb{R}$, the function

$$q(t) := t\beta^{k+1} \cdot h(t\beta), \text{ i.e., } f(t\beta) = t^k \cdot q(t) + \alpha_k t^k \beta^k + \alpha_0.$$

Since $q(0) = 0$ and the function $|q(t)|$ is continuous, there exists $t_0 \in \mathbb{R}$ with $0 < t_0 < 1$ such that

$$|q(t_0)| < |\alpha_0|.$$

We thereby obtain the estimate

$$|f(t_0\beta)| = \left|t_0^k \cdot q(t_0) - \alpha_0 t_0^k + \alpha_0\right| \leq \left|t_0^k \cdot q(t_0)\right| + |\alpha_0| \left(1 - t_0^k\right)$$
$$< t_0^k \cdot |\alpha_0| + |\alpha_0| \left(1 - t_0^k\right) = |\alpha_0| = |f(0)|.$$

But this contradicts that $|f|$ assumes its minimum at $\zeta_0 = 0$. Our assumption must therefore be false, and we must have $f(0) = 0$. That is, the polynomial $f(X)$ has at least the zero $\zeta_0 \in \mathbb{C}$. □

Remark 3.2. A field K in which the analogue of the fundamental theorem of algebra holds is said to be *algebraically closed*. Thus the field $\mathbb{C}$ of complex numbers is algebraically closed.

4. Algebraic and Transcendental Numbers

Definition 4.1. A complex number α is said to be *algebraic of degree n* if it is the root of a polynomial

$$f(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0 \tag{8}$$

of degree $n > 0$ with integer coefficients $a_0, \ldots, a_n$ but is not a root of such a polynomial of lower degree.

We denote the set of algebraic numbers by $\overline{\mathbb{Q}}$.

Remark 4.2. The set $\overline{\mathbb{Q}}$ of algebraic numbers contains all rational numbers, since every rational number $r = m/n$ ($m, n \in \mathbb{Z}; n > 0$) is algebraic of degree 1, namely as the root of the polynomial

$$f(X) = nX - m.$$

This means in addition that an algebraic number of degree $n > 1$ cannot be rational.

Example 4.3. The irrational number $\sqrt{2}$ is algebraic of degree 2, since it is a root of the polynomial $f(X) = X^2 - 2$.

Exercise 4.4. Let p be a prime number. Show that $\sqrt{p}$ is algebraic of degree 2.

Theorem 4.5. *The set $\overline{\mathbb{Q}}$ of algebraic numbers is countable.*

Proof. To prove that the set of algebraic numbers is countable, it suffices to prove that the set (8) of polynomials is countable, since each polynomial has only finitely many roots. For a fixed degree $n > 0$, there are countably many possibilities for each of the coefficients. Therefore, there are countably many polynomials of degree n with integer coefficients. Since there are only countably many choices for the degree n of a polynomial, it follows that there are countably many polynomials of positive degree with integer coefficients. This proves the theorem. □

Remark 4.6. Since the set $\mathbb{C}$ of complex numbers is uncountable, the set difference $\mathbb{T} := \mathbb{C} \setminus \overline{\mathbb{Q}}$, comprising the nonalgebraic numbers, must be uncountable. Similarly, since the set $\mathbb{R}$ of real numbers is uncountable, and the intersection $\mathbb{R} \cap \overline{\mathbb{Q}}$ is countable, the set difference $\mathbb{R} \setminus (\mathbb{R} \cap \overline{\mathbb{Q}}) = \mathbb{R} \cap \mathbb{T}$ must be uncountable.

Definition 4.7. We call a complex number $\alpha \in \mathbb{T}$ *transcendental*. A transcendental number α is thus a complex number for which there is no polynomial $f \in \mathbb{Z}[X]$ such that $f(\alpha) = 0$.

Remark 4.8. The observation made in Remark 4.6 confirms the existence of transcendental numbers. Indeed, that observation shows that transcendental numbers occur with much greater frequency than algebraic numbers. On the other hand, it does not seem particularly easy to show that a given real or complex number is indeed transcendental, since one would have to prove a negative, namely that the number is the root of no polynomial with integer coefficients. Therefore, we are familiar with many more algebraic numbers than transcendental numbers, since algebraic numbers appear (more or less) easily as roots of polynomials with integer coefficients. We shall devote the remainder of this chapter to the search for transcendental numbers. We begin with a theorem of Joseph Liouville that characterizes real algebraic numbers in terms of their approximation by rational numbers.

Theorem 4.9 (Liouville's theorem). *Let α be a real algebraic number of degree $n > 1$. Then for all $p \in \mathbb{Z}$ and sufficiently large $q \in \mathbb{N}$, we have the inequality*

$$\left| \alpha - \frac{p}{q} \right| > \frac{1}{q^{n+1}}. \tag{9}$$

This inequality says that algebraic numbers can be only "poorly" approximated by rational numbers.

Proof. Suppose the real algebraic number α is a zero of the polynomial

$$f(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0 \in \mathbb{Z}[X].$$

Moreover, let (r_m) be a sequence of rational numbers such that $\lim_{m\to\infty} r_m = \alpha$; such sequences exist, since α is real. We assume in what follows that

$$r_m = \frac{p_m}{q_m}$$

with $p_m \in \mathbb{Z}$, $q_m \in \mathbb{N}$, $q_m \neq 0$ ($m \in \mathbb{N}$). Since α is a zero of f, we have

$$\begin{aligned} f(r_m) &= f(r_m) - f(\alpha) \\ &= a_n(r_m^n - \alpha^n) + a_{n-1}(r_m^{n-1} - \alpha^{n-1}) + \cdots + a_2(r_m^2 - \alpha^2) + a_1(r_m - \alpha). \end{aligned}$$

Division by $(r_m - \alpha)$ yields

$$\frac{f(r_m)}{r_m - \alpha} = a_n(r_m^{n-1} + r_m^{n-2}\alpha + \cdots + r_m\alpha^{n-2} + \alpha^{n-1}) + \cdots + a_3(r_m^2 + r_m\alpha + \alpha^2) + a_2(r_m + \alpha) + a_1.$$

Since $\lim_{m\to\infty} r_m = \alpha$, there exists $N \in \mathbb{N}$ such that

$$|r_m - \alpha| < 1$$

for all $m \geq N$. We therefore have $|r_m| < |\alpha| + 1$ for all $m \geq N$. Using the triangle inequality, we thereby obtain for sufficiently large m the estimate

$$\left|\frac{f(r_m)}{r_m - \alpha}\right| < n \cdot |a_n| \cdot (|\alpha| + 1)^{n-1} + \cdots + 3 \cdot |a_3| \cdot (|\alpha| + 1)^2 + 2 \cdot |a_2| \cdot (|\alpha| + 1) + |a_1| =: M.$$

We note that the positive real number M is determined solely by α. In particular, it is independent of m. We now choose the index m sufficiently large that for the denominator q_m of the approximating fraction $r_m = p_m/q_m$, we have $q_m > M$. This leads to

$$\left|\frac{f(r_m)}{r_m - \alpha}\right| < q_m \quad \Longleftrightarrow \quad |\alpha - r_m| > \frac{|f(r_m)|}{q_m}. \tag{10}$$

Now, the rational numbers r_m cannot be zeros of the polynomial f, since we could then factor out the linear factor $(X - r_m)$ from f, revealing α as a zero of a polynomial of degree less than n, which cannot be the case. In other words, we have

$$|f(r_m)| = \left|\frac{a_n p_m^n + a_{n-1} p_m^{n-1} q_m + \cdots + a_1 p_m q_m^{n-1} + a_0 q_m^n}{q_m^n}\right| \neq 0. \tag{11}$$

Since the numerator in (11) is a nonzero integer, its absolute value must be at least 1. Using the estimates (10) and (11), we finally obtain

$$\left|\alpha - \frac{p_m}{q_m}\right| > \frac{|f(r_m)|}{q_m} \geq \frac{1}{q_m^n} \cdot \frac{1}{q_m} = \frac{1}{q_m^{n+1}}.$$

This completes the proof of Liouville's theorem. □

Remark 4.10. With Liouville's theorem at our disposal, we can now find transcendental numbers as follows. Assume that a given real number α is algebraic of degree $n > 0$. Showing that the inequality (9) cannot hold means that α must be transcendental. Standard examples of such numbers are real numbers whose decimal expansions contain blocks of zeros of rapidly in-

creasing length. Such numbers are called *Liouville numbers*. As an example, consider the Liouville number

$$\alpha_L := \sum_{j=1}^{\infty} 10^{-j!} = 0.110\,001\,000\,000\,000\,000\,000\,001\,000\ldots.$$

Proposition 4.11. *The Liouville number α_L is transcendental.*

Proof. For $m \in \mathbb{N}$, we set

$$p_m := 10^{m!} \cdot \sum_{j=1}^{m} 10^{-j!}, \quad q_m := 10^{m!}, \quad r_m := \frac{p_m}{q_m}.$$

We thereby obtain

$$\alpha_L - r_m = \sum_{j=1}^{\infty} 10^{-j!} - \sum_{j=1}^{m} 10^{-j!} = \sum_{j=m+1}^{\infty} 10^{-j!},$$

and it follows that

$$\begin{aligned} |\alpha_L - r_m| &= \sum_{j=m+1}^{\infty} 10^{-j!} < 10^{-(m+1)!} \cdot \sum_{j=0}^{\infty} 10^{-j} \\ &= 10^{-(m+1)!} \cdot \frac{1}{1 - \frac{1}{10}} \\ &= 10^{-(m+1)!} \cdot \frac{10}{9} < 10 \cdot 10^{-(m+1)!}. \end{aligned}$$

If α_L were algebraic of some degree n, then by Liouville's theorem, we would have, for m sufficiently large,

$$|\alpha_L - r_m| > \frac{1}{q_m^{n+1}} = \frac{1}{10^{(n+1)m!}}.$$

Combining these two inequalities yields

$$\begin{aligned} \frac{1}{10^{(n+1)m!}} < \frac{1}{10^{(m+1)!-1}} \quad &\Longleftrightarrow \quad (n+1)m! > (m+1)! - 1 \\ &\Longleftrightarrow \quad n > m - \frac{1}{m!}, \end{aligned}$$

which leads to the inequality $m < n + 1$. Since n is some fixed number, and we may choose m as large as we like, we obtain a contradiction to the assumption that α_L is algebraic. That is, α_L is transcendental. □

Exercise 4.12. Find other transcendental numbers similar to the Liouville number just discussed.

A transcendental number much better known than the Liouville numbers is Euler's number e, the base of the natural logarithm, whose transcendence we shall prove in the following section.

5. The Transcendence of *e*

Definition 5.1. *Euler's number e* is defined by the infinite series

$$\sum_{j=0}^{\infty} \frac{1}{j!}.$$

Remark 5.2. The number e thus begins with the decimal expansion $e = 2.718\,281\,828\,459\ldots$. It is the base of the *exponential function*, defined by

$$e^X := \sum_{j=0}^{\infty} \frac{X^j}{j!}.$$

For real values of X, the exponential function is strictly monotonically increasing, and it assumes only positive values. It is infinitely differentiable and is equal to all its derivatives.

Before we prove that e is transcendental, we shall begin by showing that e is irrational.

Lemma 5.3. *The number e is irrational.*

Proof. We begin by assuming for the sake of obtaining a contradiction that e is rational, that is, that $e = \frac{m}{n}$ for $m, n \in \mathbb{N}$ and $n > 0$. We now choose a natural number $k > 2$ and consider the following decomposition of the defining series for e:

$$\frac{m}{n} = e = s_k + r_k \quad \text{with} \quad s_k := \sum_{j=0}^{k} \frac{1}{j!}, \quad r_k := \sum_{j=k+1}^{\infty} \frac{1}{j!}. \tag{12}$$

We now estimate

$$\begin{aligned} r_k &= \frac{1}{(k+1)!}\left(1 + \frac{1}{k+2} + \frac{1}{(k+2)(k+3)} + \cdots\right) \\ &< \frac{1}{(k+1)!} \sum_{j=0}^{\infty} \frac{1}{(k+2)^j} = \frac{1}{(k+1)!} \cdot \frac{1}{1 - \frac{1}{k+2}} \\ &< \frac{2}{(k+1)!}. \end{aligned}$$

Multiplying (12) by $k!$, yields

$$\frac{m}{n} \cdot k! = e \cdot k! = s_k \cdot k! + r_k \cdot k!,$$

that is,

$$\frac{m}{n} \cdot k! - s_k \cdot k! = r_k \cdot k!.$$

For $k > n$, the left-hand side of this last equation represents an integer, whereas on the right-hand side, since $k > 2$, we have

$$0 < r_k \cdot k! < \frac{2k!}{(k+1)!} = \frac{2}{k+1} < 1.$$

This contradiction proves that our assumption of the rationality of e was false. □

Remark 5.4. In the following proof of the transcendence of e, we shall attempt to approximate the exponential function by a polynomial. We shall use the fact that the exponential function is characterized as the unique differentiable function $g : \mathbb{R} \longrightarrow \mathbb{R}$ satisfying the following two properties:
(1) $g'(X) = g(X) \quad (X \in \mathbb{R})$,
(2) $g(0) = 1$.
This can be seen as follows. We consider the differentiable function $e^{-X}g(X)$, whose derivative is given by

$$\left(e^{-X}g(X)\right)' = e^{-X}g'(X) - e^{-X}g(X) = 0.$$

From this, we see that the function $e^{-X}g(X)$ is constant on $\mathbb{R}$. Since $e^{-0}g(0) = 1$, this constant must be equal to 1, from which it follows that $g(X) = e^X$. For approximating the exponential function, we shall attempt to construct a polynomial whose derivative is more or less equal to the polynomial itself and whose value at the point $X = 0$ is equal to 1.

Theorem 5.5. *The number e is transcendental.*

Proof. We break the proof into six steps.

Step 1: proof strategy. We shall assume the opposite of the statement of the theorem, namely that e is algebraic of degree m, that is, that there exist $a_0, \ldots, a_m \in \mathbb{Z}$ with $a_0 \neq 0$ and $a_m \neq 0$ such that

$$a_m e^m + a_{m-1}e^{m-1} + \cdots + a_1 e + a_0 = 0.$$

We sketch in this first step how we are going to manage to obtain a contradiction. We shall assume that there exists a polynomial $H \in \mathbb{Q}[X]$ that satisfies the following four properties:
(i) $H(0) \neq 0$,

(ii) $H(j) \in \mathbb{Z} \quad (j = 0, \ldots, m)$,
(iii) $\sum_{j=0}^{m} a_j H(j) \neq 0$,
(iv) $\left|\sum_{j=1}^{m} a_j \left(H(0)e^j - H(j)\right)\right| < 1$.

In the following steps, we shall construct such a polynomial. We then set

$$c := \sum_{j=0}^{m} a_j H(j), \tag{13}$$

$$\varepsilon_j := H(0)e^j - H(j) \quad (j = 0, \ldots, m), \tag{14}$$

$$\sigma := \sum_{j=1}^{m} a_j \varepsilon_j. \tag{15}$$

Properties (ii) and (iii) show that c in (13) is a nonzero integer. Using property (i), we can transform (14) to

$$e^j = \frac{H(j)}{H(0)} + \frac{\varepsilon_j}{H(0)} \quad (j = 0, \ldots, m);$$

this can be interpreted as an approximation to the powers e^j of e $(j = 0, \ldots, m)$ by the polynomial $H(X)/H(0)$. From property (iv), we see that for σ in (15), we have $|\sigma| < 1$. Putting everything together, we have

$$\begin{aligned} 0 &= \sum_{j=0}^{m} a_j e^j \\ &= \sum_{j=0}^{m} a_j \left(\frac{H(j)}{H(0)} + \frac{\varepsilon_j}{H(0)}\right) \\ &= \frac{1}{H(0)} \sum_{j=0}^{m} a_j H(j) + \frac{1}{H(0)} \sum_{j=0}^{m} a_j \varepsilon_j \\ &= \frac{c}{H(0)} + \frac{\sigma}{H(0)}. \end{aligned}$$

After multiplying the last equation by $H(0)$ on both sides and rearranging, we end up with the equation

$$c = -\sigma, \quad \text{that is,} \quad |c| = |\sigma|. \tag{16}$$

But now we have $c \in \mathbb{Z}$ and $c \neq 0$; that is, $|c| \geq 1$. But we also have $|\sigma| < 1$. Therefore, (16) is impossible. This gives the desired contradiction to the assumption of the algebraicity of e. We conclude, then, that the number e must be transcendental.

Step 2: definition of H. We choose an arbitrary prime number p, which we shall specify more precisely later. We also define an auxiliary polynomial

$$f(X) := X^{p-1}(X-1)^p(X-2)^p \cdots (X-m)^p,$$

of degree $N = p - 1 + m \cdot p$. From it, we construct another auxiliary polynomial, namely

$$F(X) := f(X) + f'(X) + \cdots + f^{(N)}(X).$$

Since the $(N+1)$st derivative of f vanishes identically, we have

$$F'(X) = f'(X) + f''(X) + \cdots + f^{(N)}(X) = F(X) - f(X).$$

The derivative of the polynomial F would now more or less approximate F itself on the interval $[0, m]$ if f were in some sense small there. To get a handle on the size, we have to estimate the auxiliary polynomial f on the interval $[0, m]$. We observe first that

$$|X(X-1)\cdots(X-m)| \leq m^{m+1} \quad (X \in [0, m]).$$

Setting $M := m^{m+1}$ gives us the estimate

$$\max_{0 \leq X \leq m} |f(X)| \leq M^p.$$

We see that f is not small on the interval $[0, m]$. For this reason, we consider, instead of F, the polynomial

$$H(X) := \frac{F(X)}{(p-1)!}.$$

These considerations lead to the equation

$$H'(X) = H(X) - \frac{f(X)}{(p-1)!};$$

and we have also

$$\max_{0 \leq X \leq m} \left| \frac{f(X)}{(p-1)!} \right| \leq \frac{M^p}{(p-1)!}.$$

Since the quantity $M^p/(p-1)!$ can be made arbitrarily small by choosing the prime p sufficiently large, we see that the normalized polynomial $H(X)/H(0)$ approximates the exponential function e^X well on the interval $[0, m]$ if p is chosen sufficiently large.

Step 3: H satisfies property (i). We have

$$f(X) = \sum_{k=0}^{N} b_k X^k$$

with $b_0, \ldots, b_N \in \mathbb{Z}$ as well as $b_0, \ldots, b_{p-2} = 0$ and $b_{p-1} = \big((-1)^m \cdot m!\big)^p$. Since, on the other hand, we have for $k = 0, \ldots, N$ the relationship $f^{(k)}(0) = b_k \cdot k!$, we obtain

$$\begin{aligned} F(0) &= f(0) + f'(0) + \cdots + f^{(N-1)}(0) + f^{(N)}(0) \\ &= 0 + \cdots + 0 + \big((-1)^m \cdot m!\big)^p \cdot (p-1)! + b_p \cdot p! + \cdots + b_N \cdot N!, \end{aligned}$$

that is,

$$H(0) = \big((-1)^m \cdot m!\big)^p + b_p \cdot p + \cdots + b_N \cdot \frac{N!}{(p-1)!} \in \mathbb{Z}.$$

If we choose, moreover, $p > m$, then the prime p does not divide the first term in the above sum, but it does divide all the others. Therefore, we have $H(0) \neq 0$.

Step 4: H satisfies property (ii). In the previous step, we showed in particular that $H(0) \in \mathbb{Z}$. We have therefore still to prove that the property $H(j) \in \mathbb{Z}$ holds also for $j = 1, \ldots, m$. For $j = 1, \ldots, m$, we write

$$f(X) = \sum_{k=0}^{N} c_k (X - j)^k$$

with $c_0, \ldots, c_N \in \mathbb{Z}$, and we note that $c_0, \ldots, c_{p-1} = 0$, since in the definition of $f(X)$, the factor $(X - j)$ appears to the power p. On account of the relationship $f^{(k)}(j) = c_k \cdot k!$, valid for $k = 0, \ldots, N$, we may calculate

$$\begin{aligned} F(j) &= f(j) + f'(j) + \cdots + f^{(N-1)}(j) + f^{(N)}(j) \\ &= 0 + \cdots + 0 + c_p \cdot p! + \cdots + c_N \cdot N!. \end{aligned}$$

This yields

$$H(j) = c_p \cdot p + \cdots + c_N \cdot \frac{N!}{(p-1)!} \in \mathbb{Z}$$

for $j = 1, \ldots, m$, as claimed, since we have $N > p - 1$. We note here that the prime number p divides each $H(j)$ $(j = 1, \ldots, m)$.

Step 5: H satisfies property (iii). We begin by noting that property (ii) of H gives us that

$$c = \sum_{j=0}^{m} a_j H(j)$$

is an integer. Our deliberations in steps 3 and 4 now show in particular that

– $p \nmid H(0)$,
– $p \mid H(j) \quad (j = 1, \ldots, m)$.

By increasing the prime p if necessary, we may achieve as well that $p \nmid a_0 H(0)$. Then we see that

$$p \nmid \big(a_0 H(0) + a_1 H(1) + \cdots + a_m H(m)\big) \quad \Longleftrightarrow \quad p \nmid c.$$

Thus c is an integer that is not divisible by p; in particular, we have $c \neq 0$.

Step 6: H satisfies property (iv). For $t \in \mathbb{R}$, we have the differential equation

$$\frac{d}{dt}\big(F(0) - F(t)e^{-t}\big) = F(t)e^{-t} - F'(t)e^{-t} = \big(F(t) - F'(t)\big)e^{-t} = f(t)e^{-t}.$$

Applying the fundamental theorem of calculus to $X \in \mathbb{R}$ yields

$$F(0) - F(X)e^{-X} = \int_0^X f(t)e^{-t}\,dt.$$

On dividing by $(p-1)!$, we obtain at the point $X = j \in \{1, \ldots, m\}$ the equality

$$H(0) - H(j)e^{-j} = \frac{1}{(p-1)!}\int_0^j f(t)e^{-t}\,dt\,.$$

From this, we obtain the estimate

$$\begin{aligned}\Big|H(0) - H(j)e^{-j}\Big| &\leq \frac{1}{(p-1)!}\max_{0\leq X\leq m}|f(X)|\int_0^j e^{-t}\,dt \\ &\leq \frac{M^p}{(p-1)!}(1 - e^{-j}) \\ &\leq \frac{M^p}{(p-1)!}.\end{aligned}$$

On multiplying by e^j, we obtain

$$\Big|H(0)e^j - H(j)\Big| \leq \frac{M^p}{(p-1)!}e^j,$$

whence

$$\left|\sum_{j=1}^m a_j\varepsilon_j\right| = \left|\sum_{j=1}^m a_j\Big(H(0)e^j - H(j)\Big)\right| \leq \frac{M^p}{(p-1)!}\sum_{j=1}^m |a_j|e^j.$$

Since the sum $\sum_{j=1}^m |a_j|e^j$ is independent of p, and we can make the quantity $M^p/(p-1)!$ arbitrarily small by choosing p sufficiently large, we obtain for a suitable choice of the prime p the estimate

$$|\sigma| = \left|\sum_{j=1}^m a_j\varepsilon_j\right| < 1.$$

We have thereby finally shown that the polynomial H satisfies property (iv), which ensures the existence of the polynomial H having properties (i)–(iv)

postulated in the first step. This completes the proof of the transcendence of*e*. □

Remark 5.6. Even more spectacular than the proof of the transcendence of e is the proof of the transcendence of π. This result shows in particular that the number π is not constructible with straightedge and compass, which in turn proves the impossibility of squaring the circle. The proof of the transcendence of π follows along some of the same lines as that of the transcendence of e. However, at a certain point, it is necessary to bring in some of the tools of complex analysis, in particular Cauchy's integral theorem, which would take us beyond the scope of this book.

Example 5.7. We will close out the main part of this chapter by presenting two examples that illustrate how the polynomial H constructed in the proof of Theorem 5.5 can be used to obtain good approximations to the number e. Recall from that theorem that

$$H(X) = \frac{F(X)}{(p-1)!}$$

and that $H(X)/H(0) = F(X)/F(0)$ is a "good" approximation to the exponential function e^X on the interval $[0, m]$. To obtain an approximation to the number e itself, we consider the quotient $F(1)/F(0)$.

(i) We choose $m = 1$, $p = 3$ and calculate

$$\begin{aligned} f(X) &= X^2(X-1)^3, \\ F(X) &= X^5 + 2X^4 + 11X^3 + 32X^2 + 64X + 64, \\ F(0) &= 64, \quad F(1) = 174. \end{aligned}$$

We thereby obtain

$$\frac{F(1)}{F(0)} = 2.71875,$$

which is already a fairly good approximation to e.

(ii) We choose $m = 2$, $p = 5$ and calculate

$$\begin{aligned} f(X) &= X^4(X-1)^5(X-2)^5, \\ F(X) &= X^{14} - X^{13} + 87X^{12} + 654X^{11} + \cdots + 29\,141\,344\,128, \\ F(0) &= 29\,141\,344\,128, \quad F(1) = 79\,214\,386\,200. \end{aligned}$$

We now obtain

$$\frac{F(1)}{F(0)} = 2.718\,281\,828\,458\,561\ldots,$$

which agrees with the decimal expansion of e to ten decimal places.

Exercise 5.8. Compute in the way described above further approximations to e.

E. Zeros of Polynomials: The Search for Solution Formulas

The point of departure for this appendix is the fundamental theorem of algebra, which we proved in this chapter as Theorem 3.1. The theorem says that every polynomial $f(X)$ of positive degree n with complex coefficients, that is, $f \in \mathbb{C}[X]$, has all of its zeros in $\mathbb{C}$.

Knowing as we do that the zeros of every quadratic polynomial can be expressed through an explicit formula involving the four arithmetic operations and the extraction of roots, the question naturally arises whether similar formulas exist for polynomials of higher degree. It is this question and its consequences that we wish to explore in this appendix. And as we do, we shall come to realize that efforts to resolve this problem extend to active research in number theory today.

E.1 Zeros of Polynomials of Degree $n \leq 4$

The zeros of linear and quadratic polynomials with complex coefficients are easily determined. In the quadratic case, we presented the solution formulas in the proof of Theorem 1.7, in connection with which the reader should also take note of Remark 1.9.

We now turn to determining the zeros of an arbitrary third-degree polynomial $f(X)$ with complex coefficients. First of all, we may, without loss of generality, assume that $f(X)$ has the form

$$f(X) = X^3 + \beta X + \gamma \tag{17}$$

with $\beta, \gamma \in \mathbb{C}$. Namely, if $f(X) = X^3 + \alpha' X^2 + \beta' X + \gamma'$, then one can obtain the desired form (17) by means of the substitution $X \mapsto X - \alpha'/3$, known as the *Tschirnhaus transformation*.

If we now decompose a zero $\zeta \in \mathbb{C}$ of the polynomial (17) as $\zeta = \xi + \eta$, we obtain the equation

$$3\xi\eta\zeta + \xi^3 + \eta^3 = \zeta^3 = -\beta\zeta - \gamma.$$

Comparing coefficients yields

$$\xi^3 + \eta^3 = -\gamma \quad \text{and} \quad \xi \cdot \eta = -\frac{\beta}{3}, \quad \text{whence} \quad \xi^3 \cdot \eta^3 = -\left(\frac{\beta}{3}\right)^3$$

and thus ξ^3 and η^3 can be viewed as the two zeros of the quadratic polynomial

$$X^2 + \gamma X - \frac{\beta^3}{27}.$$

This is, incidentally, the assertion of Viète's theorem. This polynomial is called the *quadratic resolvent* of the cubic polynomial (17). The first zero of (17) thus takes the form

$$\zeta_1 = \sqrt[3]{-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}} + \sqrt[3]{-\frac{\gamma}{2} - \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}}.$$

The other two zeros of the cubic polynomial (17) can be obtained with the help of a complex third root of unity ε, that is, a complex number ε satisfying $\varepsilon^3 = 1$, for example,

$$\varepsilon = -\frac{1}{2} + \frac{\sqrt{3}}{2} i.$$

Taking into account the relation $\xi\eta = -\beta/3$, we obtain

$$\zeta_2 = \varepsilon \sqrt[3]{-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}} + \varepsilon^2 \sqrt[3]{-\frac{\gamma}{2} - \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}},$$

$$\zeta_3 = \varepsilon^2 \sqrt[3]{-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}} + \varepsilon \sqrt[3]{-\frac{\gamma}{2} - \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}}.$$

These solution formulas first appeared in Girolamo Cardano's 1545 book *Ars Magna* and are therefore known as *Cardano's formulas*; they had been discovered earlier by Niccolò Tartaglia. Altogether, we may state that in the case of a cubic polynomial, its zeros can be expressed in terms of (square and cube) roots of the polynomial's coefficients.

We now turn our attention to determining the zeros of an arbitrary fourth-degree polynomial $f(X)$ with complex coefficients. In analogy to the cubic case, we may assume without loss of generality that $f(X)$ is of the form

$$f(X) = X^4 + \beta X^2 + \gamma X + \delta, \tag{18}$$

with $\beta, \gamma, \delta \in \mathbb{C}$. As in the previous case, the problem of finding the zeros of (18) will be reduced to determining the zeros of a polynomial of lower degree, the so-called *cubic resolvent*, which is given by the cubic polynomial

$$X^3 + 2\beta X^2 + (\beta^2 - 4\delta)X - \gamma^2. \tag{19}$$

If we denote the three zeros of the cubic resolvent (19) by η_1, η_2, η_3, the four zeros of (18) can be represented in the form

$$\zeta_1 = \frac{+\sqrt{\eta_1} + \sqrt{\eta_2} + \sqrt{\eta_3}}{2}, \qquad \zeta_2 = \frac{+\sqrt{\eta_1} - \sqrt{\eta_2} - \sqrt{\eta_3}}{2},$$
$$\zeta_3 = \frac{-\sqrt{\eta_1} + \sqrt{\eta_2} - \sqrt{\eta_3}}{2}, \qquad \zeta_4 = \frac{-\sqrt{\eta_1} - \sqrt{\eta_2} + \sqrt{\eta_3}}{2}.$$

It thus turns out that in this case as well, the zeros that we are seeking can be expressed in terms of roots of the coefficients of the underlying quartic polynomial. The solution formulas presented here also first appeared in Cardano's book *Ars Magna*; they were first discovered by Ludovico Ferrari.

E.2 Zeros of Polynomials of Degree $n = 5$

The description of the zeros of polynomials of degree $n \leq 4$ readily leads to the conjecture that the zeros of higher-degree polynomials can also be reduced, with the help of expressions involving the extraction of roots, to the determination of the zeros of polynomials of lower degree, whose zeros, in turn, can also be expressed in terms of expressions involving the extraction of roots. We shall see that in general, this conjecture is false, as was proved by Niels Henrik Abel at the beginning of the nineteenth century. In order to present Abel's results, we begin by introducing some general concepts.

Let $f \in \mathbb{C}[X]$ be a polynomial of degree $n > 0$, which we shall write in the form

$$f(X) = \beta_n X^n - \beta_{n-1} X^{n-1} \pm \cdots \pm \beta_1 X + (-1)^n \beta_0 ,$$

with $\beta_0, \ldots, \beta_n \in \mathbb{C}$ and $\beta_n \neq 0$. For the sake of simplicity, in what follows we shall assume that f is monic, that is, that $\beta_n = 1$. If we denote the zeros of f by $\zeta_1, \ldots, \zeta_n$, we obtain the factorization

$$f(X) = (X - \zeta_1) \cdots (X - \zeta_n).$$

By introducing along with the indeterminate X the additional independent indeterminates $X_1, \ldots, X_n$, we define the *general nth-degree polynomial* by the formula

$$f_{\text{gen}}(X) := (X - X_1) \cdots (X - X_n),$$

which on multiplying out the linear factors takes the form

$$f_{\text{gen}}(X) = X^n - \sigma_1 X^{n-1} \pm \cdots \pm \sigma_{n-1} X + (-1)^n \sigma_n ,$$

where the coefficients $\sigma_1,\dots,\sigma_n$ are given by the *elementary symmetric polynomials*

$$\begin{aligned}\sigma_1 &= \sigma_1(X_1,\dots,X_n) = \sum_{j=1}^{n} X_j,\\ \sigma_2 &= \sigma_2(X_1,\dots,X_n) = \sum_{\substack{j,k=1\\ j<k}}^{n} X_jX_k,\\ &\dots\\ \sigma_n &= \sigma_n(X_1,\dots,X_n) = X_1\cdots X_n.\end{aligned}$$

The coefficients of f_{gen} lie in the field of rational functions in the indeterminates $\sigma_1,\dots,\sigma_n$, that is, in the quotient field $\operatorname{Quot}(\mathbb{C}[\sigma_1,\dots,\sigma_n])$ of the polynomial ring $\mathbb{C}[\sigma_1,\dots,\sigma_n]$, whose elements are quotients of polynomials in the indeterminates $\sigma_1,\dots,\sigma_n$ and which we denote by $\mathbb{C}(\sigma_1,\dots,\sigma_n)$. The zeros of the polynomial

$$f_{\text{gen}} \in \mathbb{C}(\sigma_1,\dots,\sigma_n)[X]$$

lie in the field of rational functions in the indeterminates $X_1,\dots,X_n$, that is, in the field $\mathbb{C}(X_1,\dots,X_n)$, which contains the field $\mathbb{C}(\sigma_1,\dots,\sigma_n)$.

Example E.1. The general polynomial of degree 2 is given by

$$\begin{aligned}f_{\text{gen}}(X) &= (X-X_1)\cdot(X-X_2)\\ &= X^2-(X_1+X_2)X+X_1X_2\\ &= X^2-\sigma_1X+\sigma_2.\end{aligned}$$

Its coefficients are the elementary symmetric polynomials $\sigma_1=\sigma_1(X_1,X_2)=X_1+X_2$ and $\sigma_2=\sigma_2(X_1,X_2)=X_1X_2$. We therefore have $f_{\text{gen}}\in\mathbb{C}(\sigma_1,\sigma_2)[X]$. The two zeros X_1 and X_2 of $f_{\text{gen}}(X)$ are elements of the field $\mathbb{C}(X_1,X_2)$. By specializing σ_1 and σ_2, that is, by evaluating σ_1 and σ_2 at particular complex numbers, one can obtain every monic second-degree polynomial $f\in\mathbb{C}[X]$. This explains why we call $f_{\text{gen}}(X)$ the general polynomial of degree 2.

Definition E.2. Let $f_{\text{gen}}\in\mathbb{C}(\sigma_1,\dots,\sigma_n)[X]$ be the general polynomial of degree n. We say that the zeros $X_1,\dots,X_n$ *can be expressed in terms of radicals* if there exist $m\in\mathbb{N}_{>0}$ and polynomials $p_0,p_1,\dots,p_{m-1},R\in\mathbb{C}(\sigma_1,\dots,\sigma_n)$ with $R^{1/m}\notin\mathbb{C}(\sigma_1,\dots,\sigma_n)$ such that every X_j can be represented in the form

$$X_j = p_0+p_1R^{1/m}+p_2R^{2/m}+\cdots+p_{m-1}R^{(m-1)/m} \tag{20}$$

or more generally as a finite iteration of such expressions. Here the dependence of the right-hand side of (20) on the index j ($j=1,\dots,n$) comes into

play, in that various choices of the radical $R^{1/m}$ can be made, which differ from one another by an mth root of unity.

Example E.3. For the two zeros X_1, X_2 of the general quadratic polynomial $f_{\text{gen}}(X) = X^2 - \sigma_1 X + \sigma_2$, one has

$$X_{1,2} = \frac{\sigma_1 \pm \sqrt{\sigma_1^2 - 4\sigma_2}}{2} = \frac{\sigma_1}{2} \pm \frac{1}{2} R^{1/2},$$

with $p_0 = \sigma_1/2$, $p_1 = 1/2$, and $R = \sigma_1^2 - 4\sigma_2 \in \mathbb{C}(\sigma_1, \sigma_2)$. Thus the two zeros X_1, X_2 of $f_{\text{gen}}(X)$ can be expressed in terms of radicals.

Similarly, one can easily see, using the solution formulas given above, that the zeros of the general cubic and quartic polynomials can also be expressed in terms of radicals. However, one has the following theorem.

Theorem E.4 (Abel). *Let $f_{\text{gen}} \in \mathbb{C}(\sigma_1, \ldots, \sigma_5)[X]$ denote the general polynomial of degree 5. Then the zeros $X_1, \ldots, X_5$ of $f_{\text{gen}}(X)$ cannot be represented in terms of radicals.*

Proof. We shall briefly sketch the idea of the proof. We begin by supposing, in contradiction to the assertion of the theorem, that the zeros of f_{gen} are indeed representable in terms of radicals. Based on that assumption, it turns out that the zeros $X_1, \ldots, X_5$ must satisfy an algebraic relation over $\mathbb{C}$, which contradicts the assumption that f_{gen} is the *general* polynomial of degree 5, whose zeros are therefore algebraically independent over $\mathbb{C}$ and therefore can satisfy no polynomial relation over $\mathbb{C}$. □

E.3 The Bridge to Group Theory: Galois Theory

The negative result given by Abel's theorem, Theorem E.4, raises the problem of providing a conceptual characterization of the zeros of polynomials whose coefficients lie in a fixed field K. With this in mind, we again consider the general polynomial $f_{\text{gen}} \in \mathbb{C}(\sigma_1, \ldots, \sigma_n)[X]$ of degree n and observe that $K := \mathbb{C}(\sigma_1, \ldots, \sigma_n)$ can be characterized by the fact that K is the field constructed of all rational functions from $E := \mathbb{C}(X_1, \ldots, X_n)$ that remain invariant under permutations of the indeterminates $X_1, \ldots, X_n$. This insight arises from the nontrivial fact that a polynomial $g \in \mathbb{C}[X_1, \ldots, X_n]$ that remains invariant under all permutations of the indeterminates $X_1, \ldots, X_n$ must be a polynomial in the elementary symmetric polynomials $\sigma_1, \ldots, \sigma_n$. This result carries over directly to rational functions in the indeterminates $X_1, \ldots, X_n$ that are invariant under all permutations of $X_1, \ldots, X_n$. We have thereby, in a natural way, associated with the field E and subfield K a characterizing group, the nth symmetric group S_n (see Example 2.8 (iv) of Chapter II). This

association is the starting point of Galois theory, whose basic features we shall now present.

Definition E.5. Let K be an arbitrary field. A field E that contains the field K is called a *field extension of* K. The extension $E \supseteq K$ is denoted by E/K, and we say "E over K." A field extension E of K can be viewed naturally as a K-vector space, and such a field extension is said to be *finite over* K if E is finite-dimensional as a K-vector space. We let $[E : K]$ denote the dimension $\dim_K E$ and call this number the *degree of* E *over* K.

Definition E.6. Let K be an arbitrary field, and E/K a field extension. We say that an element $\alpha \in E$ is *algebraic over* K if α is a zero of a polynomial $f \in K[X]$. A field extension E of K is said to be *algebraic over* K if all of its elements are algebraic over K; we also say that we have an *algebraic field extension* E/K.

Example E.7. For $K = \mathbb{Q}$ and $E = \mathbb{C}$, we saw in Section 4 of this chapter that a number $\alpha \in \mathbb{C}$ is algebraic over $\mathbb{Q}$ precisely when $\alpha \in \overline{\mathbb{Q}}$, that is, when α is an algebraic number. In particular, transcendental numbers are not algebraic over $\mathbb{Q}$. We have therefore that neither $\mathbb{C}/\mathbb{Q}$ nor $\mathbb{R}/\mathbb{Q}$ is an algebraic field extension. We saw, however, in Section 1 of this chapter that $\mathbb{C}/\mathbb{R}$ is an algebraic and finite field extension of degree $[\mathbb{C} : \mathbb{R}] = \dim_{\mathbb{R}} \mathbb{C} = 2$.

Remark E.8. One can see at once that a finite field extension E/K is algebraic, since the $[E : K] + 1$ powers

$$1,\ \alpha,\ \alpha^2,\ \ldots,\ \alpha^{[E:K]}$$

of an arbitrary element $\alpha \in E$ must be linearly dependent over K, whence α is a zero of a polynomial with coefficients in K.

Remark E.9. If α is a zero of a nontrivial polynomial in the polynomial ring $K[X]$, then there exists a monic polynomial f of minimal degree with α as a zero. Such a polynomial f is uniquely determined and is called the minimal polynomial of α. The existence of such a polynomial f can be seen simply by considering the set

$$\mathfrak{a}_\alpha := \{g \in K[X] \mid g(\alpha) = 0\},$$

which is obviously a nontrivial ideal of $K[X]$, and indeed, it is a principal ideal, since $K[X]$ is a Euclidean domain. That is, $\mathfrak{a}_\alpha = (f)$ with f a polynomial of minimal positive degree, which can be made monic. The uniqueness can be seen at once by applying the Euclidean algorithm. Moreover, one sees that the minimal polynomial f of α is irreducible over K.

Example E.10. Let $K = \mathbb{Q}$ and

$$E = \{\alpha = a + b\sqrt{-3} \mid a, b \in \mathbb{Q}\}.$$

It is easily verified that E is a field that contains the field $K = \mathbb{Q}$. The field extension E/K is finite over K. Indeed, we have $[E : K] = 2$, since the elements 1 and $\sqrt{-3}$ constitute a basis of E over K. And so E is also algebraic over K. The minimal polynomial of the element $\alpha = \sqrt{-3}$ is $f(X) = X^2 + 3$.

Remark E.11. In general, we can construct a field extension of K that contains the zero α of a polynomial f that is irreducible over K. To this end, we consider the ring homomorphism $\varphi : K[X] \to K[\alpha]$ given by replacing the indeterminate X with the element α. Since the kernel of the homomorphism φ is the ideal $\mathfrak{a}_\alpha = (f)$, the homomorphism theorem for rings gives us the isomorphism

$$K[\alpha] \cong K[X]/(f).$$

Since the polynomial f is irreducible over K, the principal ideal (f) is a prime ideal, indeed a maximal ideal, which means that the quotient ring $K[X]/(f)$ is a field. This field is the desired extension of K that contains the element α. Moreover, our construction shows that the ring $K[\alpha]$ is in fact a field, equal to its field of quotients $K(\alpha)$. It is easy to see that $K(\alpha)$ is a finite field extension of K and that $[K(\alpha) : K] = \deg(f)$. A basis of $K(\alpha)$ over K is given by the elements

$$1,\ \alpha,\ \alpha^2,\ \ldots,\ \alpha^{\deg(f)-1}.$$

One says that the field $K(\alpha)$ is constructed by *adjoining α to K*.

Example E.12. From these observations, we obtain for Example E.10 that $E = \mathbb{Q}(\sqrt{-3})$. In particular, we conclude that $\mathbb{Q} \subset E \subset \overline{\mathbb{Q}}$. By considering additional square roots, one sees that there are infinitely many distinct fields lying between $\mathbb{Q}$ and $\overline{\mathbb{Q}}$ that are algebraic over $\mathbb{Q}$.

Remark E.13. Every finite field extension E/K can be constructed by successively adjoining finitely many elements $\alpha_1, \ldots, \alpha_n$ that are algebraic over K. One thereby obtains E in the form

$$E = K(\alpha_1)(\alpha_2)\cdots(\alpha_n) =: K(\alpha_1, \ldots, \alpha_n).$$

We shall now restrict our attention to *finite* field extensions E/K, and shall further assume that $\mathrm{char}(K) = 0$.

Definition E.14. Let E/K be a finite field extension. A *K-isomorphism of E* is a field isomorphism of E into (some) field E'/K that leaves each element of K fixed. We denote the set of K-isomorphisms of E (into some field E'/K) by $\mathrm{Iso}_K(E)$.

The subset of K-automorphisms of E is denoted by $\mathrm{Aut}_K(E)$. This set is clearly a group. If we have the equality $\mathrm{Aut}_K(E) = \mathrm{Iso}_K(E)$, then E is called a *Galois extension of* K. The group

$$\mathrm{Gal}(E/K) := \mathrm{Aut}_K(E)$$

is called the *Galois group of* E/K.

Example E.15. For example, if α and α' are zeros of the irreducible polynomial $f \in K[X]$, then the assignment $\alpha \mapsto \alpha'$ induces a K-isomorphism of $K(\alpha)$ to $K(\alpha')$.

Remark E.16. If E/K is a Galois extension of K, then clearly, $|\mathrm{Gal}(E/K)| = [E : K]$.

Example E.17. The finite field extension $\mathbb{Q}(\sqrt{-3})/\mathbb{Q}$ from Example E.10 is a Galois extension with Galois group $\mathrm{Gal}(\mathbb{Q}(\sqrt{-3})/\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z}$. The two $\mathbb{Q}$-automorphisms of $\mathbb{Q}(\sqrt{-3})$ are given by the assignments

$$\mathrm{id} : a + b\sqrt{-3} \mapsto a + b\sqrt{-3} \quad \text{and} \quad \sigma : a + b\sqrt{-3} \mapsto a - b\sqrt{-3}.$$

Example E.18. Let $\alpha = \sqrt[3]{2} \in \mathbb{R}$ be the real zero of $X^3 - 2$. The other two zeros of $X^3 - 2$ are not real, being given by $\zeta\alpha$ and $\zeta^2\alpha$, with $\zeta = e^{2\pi i/3}$. The field extension $\mathbb{Q}(\alpha)/\mathbb{Q}$ is finite of degree $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, but it is not a Galois extension. Namely, if $\varphi \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\alpha))$ is an arbitrary $\mathbb{Q}$-automorphism, then $\varphi(\alpha)$ is a zero of the polynomial $X^3 - 2$, yet we must have $\varphi(\alpha) \in \mathbb{Q}(\alpha) \subset \mathbb{R}$. This shows that $\varphi(\alpha) = \alpha$ and therefore $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\alpha)) = \{\mathrm{id}\}$; that is, we have $|\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\alpha))| = 1$. Taking Remark E.16 into account, we see that $\mathbb{Q}(\alpha)/\mathbb{Q}$ cannot be a Galois extension.

The fundamental theorem of Galois theory establishes the following correspondence between fields and groups.

Theorem E.19 (Fundamental theorem of Galois theory). *With the foregoing notation and assumptions, let E/K be a Galois extension with Galois group $\mathrm{Gal}(E/K)$. We consider the sets*

$$\mathcal{K} := \{L \text{ a field} \mid K \subseteq L \subseteq E\},$$
$$\mathcal{G} := \{H \text{ a group} \mid \{\mathrm{id}\} \leq H \leq \mathrm{Gal}(E/K)\}.$$

Then there exists a one-to-one correspondence between the sets $\mathcal{K}$ and $\mathcal{G}$.

Proof. In the following proof sketch, we shall present the mutually inverse mappings of $\mathcal{K}$ to $\mathcal{G}$ and from $\mathcal{G}$ to $\mathcal{K}$ without providing a rigorous proof of bijectivity. To this end, we set $G := \mathrm{Gal}(E/K)$. The mapping

$$\varphi : \mathcal{K} \longrightarrow \mathcal{G}$$

is given by the assignment

$$L \mapsto G^L := \{g \in G \mid g(\alpha) = \alpha, \forall \alpha \in L\};$$

it is easily verified that the set G^L is in fact a group, and therefore belongs to $\mathcal{G}$. The inverse mapping

$$\psi : \mathcal{G} \longrightarrow \mathcal{K}$$

is given by the assignment

$$H \mapsto E^H := \{\alpha \in E \mid g(\alpha) = \alpha, \forall g \in H\},$$

and again one easily checks that the set E^H is a field with $K \subseteq E^H$, and therefore belongs to $\mathcal{K}$.

As we mentioned, the proof consists in showing that the two mappings φ and ψ are inverses of each other. □

Remark E.20. The fundamental theorem of Galois theory, Theorem E.19, shows in particular that under the above assumptions, E is a Galois extension of every intermediate field $K \subseteq L \subseteq E$ with Galois group $\mathrm{Gal}(E/L) = \mathrm{Gal}(E/K)^L$.

Example E.21. Let $\alpha = \sqrt[3]{2}$ and $\zeta = e^{2\pi i/3}$ be as in Example E.18, $K = \mathbb{Q}$, and $E = \mathbb{Q}\big(\alpha, \zeta\alpha, \zeta^2\alpha\big) = \mathbb{Q}(\alpha, \zeta)$. The field E is the smallest field that contains all the zeros of the polynomial $f(X) = X^3 - 2$ and is therefore, by Definition E.24 below, a Galois extension of $\mathbb{Q}$. The Galois group $\mathrm{Gal}(E/\mathbb{Q})$ consists of the $\mathbb{Q}$-automorphisms induced by all the permutations of the three zeros $\alpha_1 := \alpha$, $\alpha_2 := \zeta\alpha$, $\alpha_3 := \zeta^2\alpha$ of the polynomial $f(X)$. Since these six permutations lead to six distinct $\mathbb{Q}$-automorphisms of E, we conclude that the Galois group of $E/\mathbb{Q}$ is the symmetric group S_3. With the notation of Example 4.23 of Chapter II, we thereby obtain

$$\mathrm{Gal}(E/\mathbb{Q}) = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\},$$

where the action of π_j ($j = 1, \ldots, 6$) on the zeros $\alpha_1, \alpha_2, \alpha_3$ is described by the corresponding permutation of the indices.

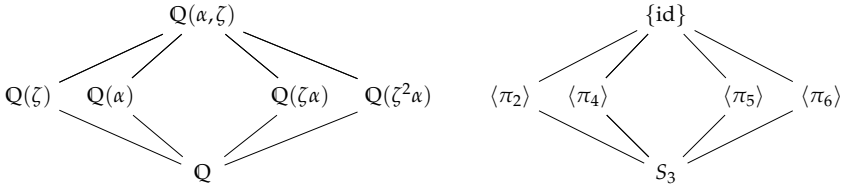
In Exercise 2.26 of Chapter II, we showed that S_3 has six subgroups: the group S_3 itself and the following five proper subgroups:

$$\langle \pi_1 \rangle = \{\mathrm{id}\},\ \langle \pi_2 \rangle = \langle \pi_3 \rangle = \{\pi_1, \pi_2, \pi_3\},$$
$$\langle \pi_4 \rangle = \{\pi_1, \pi_4\},\ \langle \pi_5 \rangle = \{\pi_1, \pi_5\},\ \langle \pi_6 \rangle = \{\pi_1, \pi_6\}.$$

We therefore have

$$\mathcal{G} = \big\{\{\mathrm{id}\}, \langle\pi_2\rangle, \langle\pi_4\rangle, \langle\pi_5\rangle, \langle\pi_6\rangle, S_3\big\}.$$

By the fundamental theorem of Galois theory, the trivial subgroup $\{\mathrm{id}\}$ corresponds to the field $\mathbb{Q}(\alpha,\zeta)$, while the group S_3 corresponds to the field $\mathbb{Q}$. The field extension $\mathbb{Q}(\alpha,\zeta)/\mathbb{Q}$ must therefore have precisely four strictly intermediate fields. They are $\mathbb{Q}(\zeta)$, $\mathbb{Q}(\alpha)$, $\mathbb{Q}(\zeta\alpha)$, $\mathbb{Q}(\zeta^2\alpha)$, and they correspond to the four remaining subgroups, as shown in the following diagram, where the intermediate fields are shown on the left, and their corresponding subgroups on the right at the corresponding locations.



E.4 Zeros of Polynomials and Galois Theory

We begin with the following definition, which brings us back to group theory.

Definition E.22. Let G be a finite group with identity element e. A *normal series of* G is a finite descending chain of subgroups

$$G = G_0 \geq G_1 \geq \cdots \geq G_{n-1} \geq G_n = \{e\} \tag{21}$$

such that for $j = 1,\ldots,n$, the subgroup G_j is a normal subgroup of G_{j-1}.

Furthermore, a group G is said to be *solvable* if it possesses a normal series of the form (21) such that the quotient groups G_{j-1}/G_j for $j = 1,\ldots,n$ are abelian.

Example E.23. The symmetric group S_4 of permutations on four symbols is solvable, for it possesses the normal series

$$S_4 > A_4 > V_4 > U_2 > \{\mathrm{id}\},$$

where A_4 is the alternating group on four symbols,

$$V_4 = \{\mathrm{id}, (12)(34), (13)(24), (14)(23)\}$$

is a subgroup of order 4 isomorphic to the dihedral group D_4 from Example 2.8 (iii) of Chapter II, and $U_2 = \{\mathrm{id}, (12)(34)\}$ is a subgroup of order 2. Here the notation (jk) $(j,k = 1,\ldots,4;\, j \neq k)$ represents the permutation (transposition) in S_4 that interchanges j and k while leaving all the other elements

fixed. For the successive quotient groups, we have the group isomorphisms

$$S_4/A_4 \cong \mathbb{Z}/2\mathbb{Z}, \quad A_4/V_4 \cong \mathbb{Z}/3\mathbb{Z}, \quad V_4/U_2 \cong \mathbb{Z}/2\mathbb{Z}, \quad U_2/\{\mathrm{id}\} \cong \mathbb{Z}/2\mathbb{Z}.$$

A similar analysis shows that the symmetric groups S_n for $n = 1,2,3$ are solvable. In contrast, it can be shown that the symmetric groups S_n for $n \geq 5$ are *not* solvable.

Definition E.24. Let $f \in K[X]$ be a polynomial of degree n with zeros $\zeta_1,\ldots,\zeta_n$. Then the smallest field E that contains all these zeros is given by the finite field extension $E = K(\zeta_1,\ldots,\zeta_n)$, called the *splitting field of* f. The splitting field E is a Galois extension of K. We define the *Galois group* $\mathrm{Gal}(f)$ *of* f to be the Galois group $\mathrm{Gal}(E/K)$.

Example E.25. The general polynomial $f_{\mathrm{gen}} \in \mathbb{C}(\sigma_1,\ldots,\sigma_n)[X]$ of degree n has the splitting field $\mathbb{C}(X_1,\ldots,X_n)$. The Galois group of f_{gen} thus consists of all $\mathbb{C}(\sigma_1,\ldots,\sigma_n)$-automorphisms that permute the n zeros $X_1,\ldots,X_n$ of the polynomial $f_{\mathrm{gen}}(X)$ in all possible ways. That is, $\mathrm{Gal}(f_{\mathrm{gen}}) = S_n$.

The following theorem concludes our discussion of the representation of the zeros of a polynomial $f \in K[X]$ by radicals by providing a complete characterization of the Galois group $\mathrm{Gal}(f)$ of f in terms of its group-theoretic properties.

Theorem E.26 (Representation of zeros by radicals). *The zeros of a polynomial $f \in K[X]$ are representable by radicals if and only if* $\mathrm{Gal}(f)$*, the Galois group of f, is solvable as a group.* □

Remark E.27. Taking into account Examples E.23 and E.25, Theorem E.26 gives a new proof of Abel's theorem.

E.5 A Way Out of a Dilemma: The Case of the Ground Field $\mathbb{Q}$

As a consequence of the negative result of Theorems E.4 and E.26, the question arises how we might somehow conceptually "get a handle" on the zeros of a polynomial $f \in K[X]$. The answer to this question is one of the central tasks of algebraic number theory. In this last section, we would like to provide an answer to this question in the case $K = \mathbb{Q}$.

Since we know by Theorem E.26 that the zeros $\zeta_1,\ldots,\zeta_n$ of a polynomial $f \in \mathbb{Q}[X]$ of degree $n \geq 5$ cannot in general be represented in terms of radicals and that therefore, the splitting field $E = \mathbb{Q}(\zeta_1,\ldots,\zeta_n)$ is not easily accessible, it is clear that we need to find another way to pursue our investigation of the Galois group $\mathrm{Gal}(f) = \mathrm{Gal}(E/\mathbb{Q})$. Before we can do this, we need some facts about the arithmetic properties underlying the field E in order

to characterize the Galois group $\mathrm{Gal}(E/\mathbb{Q})$. This theory goes back to David Hilbert. For this, we draw on the first chapter of the book [8].

If f is a polynomial of degree $n = 1$, then $E = \mathbb{Q}$, and the arithmetic of the field E is described by the ring of integers $\mathbb{Z}$, which was our object of study in Sections 2, 3, and 4 of Chapter I. Once the degree of the polynomial n is greater than 1, we shall need an analogue in the field E of the ring of integers $\mathbb{Z}$, whose arithmetic we shall be able to describe with the help of the ideal theory developed in Section 7 of Chapter III. In particular, the divisibility of ideals and the notions of prime ideal and maximal ideal will be our focus. For this, we collect in the following definition some terminology and facts.

Definition E.28. The *ring of integers $\mathcal{O}_E$ of a field extension $E/\mathbb{Q}$* consists of all $\alpha \in E$ that are zeros of a *monic* polynomial with *rational integer* coefficients, that is, with coefficients that are in $\mathbb{Z}$.

Remark E.29. The ring of integers $\mathcal{O}_E$ of the field extension $E/\mathbb{Q}$ is a commutative ring contained in the field E and containing the ring of rational integers $\mathbb{Z}$. In analogy to the fundamental theorem of arithmetic, Theorem 3.1 of Chapter I, it turns out that on the level of ideals of $\mathcal{O}_E$, every ideal $\mathfrak{a} \subseteq \mathcal{O}_E$ can be uniquely (up to order) represented as a product of positive powers of prime ideals. That is,

$$\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_r^{a_r}$$

for some $r \in \mathbb{N}$ and distinct prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ with positive integer exponents $a_1, \ldots, a_r$.

Definition E.30. If p is a prime number, then the principal ideal (p) is an ideal of the ring of integers $\mathcal{O}_E$ with a prime ideal decomposition of the form

$$(p) = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}. \tag{22}$$

The exponent e_j is called the *ramification index of $\mathfrak{p}_j$ over p* ($j = 1, \ldots, r$).

Since the prime ideals $\mathfrak{p}_j$ must be maximal ideals (otherwise, they could be further decomposed), the quotient rings $\mathcal{O}_E/\mathfrak{p}_j$ are fields, which are clearly finite field extensions of the field $\mathbb{F}_p$ with p elements. We set $f_j := [\mathcal{O}_E/\mathfrak{p}_j : \mathbb{F}_p]$ and call it the *residue class degree of $\mathfrak{p}_j$ over p* ($j = 1, \ldots, r$).

Remark E.31. It turns out that among the quantities we have just defined, there exists the following fundamental relation:

$$\sum_{j=1}^{r} e_j f_j = [E : \mathbb{Q}].$$

In the present case, this reduces to the equality $e \cdot f \cdot r = [E : \mathbb{Q}]$, since E is a Galois extension of $\mathbb{Q}$, with the result that all the ramification indices and

all the residue class degrees are equal; that is, we have $e_j = e$ and $f_j = f$ for $j = 1, \ldots, r$. The quantities e, f, r therefore depend solely on the prime number p. The decomposition (22) can thus be simplified to

$$(p) = (\mathfrak{p}_1 \cdots \mathfrak{p}_r)^e. \tag{23}$$

Definition E.32. If for the Galois extension $E/\mathbb{Q}$ and prime number p there exists a prime ideal decomposition (23) with $e = 1$, then p is said to be *unramified in E*.

Remark E.33. Since the splitting field $E = \mathbb{Q}(\zeta_1, \ldots, \zeta_n)$ of f is a finite Galois extension of $\mathbb{Q}$, it follows that the $\mathcal{O}_E/\mathfrak{p}_j$ are also finite Galois extensions of $\mathbb{F}_p$. The Galois groups $\mathrm{Gal}(\mathcal{O}_E/\mathfrak{p}_j/\mathbb{F}_p)$ turn out in the unramified case (that is, $e = 1$) to be isomorphic to the subgroups

$$D_j := \{\sigma \in \mathrm{Gal}(E/\mathbb{Q}) \mid \sigma(\alpha) \equiv \alpha \mod \mathfrak{p}_j, \forall \alpha \in E\}, \quad j = 1, \ldots, r,$$

of $\mathrm{Gal}(E/\mathbb{Q})$, which are called the *decomposition groups of* $\mathfrak{p}_j$ and are all conjugate to one another.

Since the finite fields $\mathcal{O}_E/\mathfrak{p}_j$ are finite Galois extensions of $\mathbb{F}_p$, the Galois groups $\mathrm{Gal}(\mathcal{O}_E/\mathfrak{p}_j/\mathbb{F}_p)$ are cyclic. Each is generated by the automorphism that takes the residue class $\bar{\alpha} = \alpha \mod \mathfrak{p}_j \in \mathcal{O}_E/\mathfrak{p}_j$ to $\bar{\alpha}^p$.

Definition E.34. The automorphism of $\mathrm{Gal}(E/\mathbb{Q})$ just defined in the unramified case is called the *Frobenius automorphism for* $\mathfrak{p}_j$ and denoted by $\mathrm{Frob}_{\mathfrak{p}_j}$. The Frobenius automorphisms $\mathrm{Frob}_{\mathfrak{p}_j}$, for $j = 1, \ldots, r$, are conjugate to one another. The conjugation class depends only on the prime number p.

Remark E.35. If p is a prime that is unramified in E, this corresponds to a cyclic subgroup D_j of the Galois group $\mathrm{Gal}(E/\mathbb{Q})$ that is generated by the Frobenius automorphism $\mathrm{Frob}_{\mathfrak{p}_j}$ and is uniquely determined up to conjugation. By the fundamental theorem of Galois theory, Theorem E.19, associated with the subgroup D_j is an intermediate field $\mathbb{Q} \subseteq L_j \subseteq E$ such that $\mathrm{Gal}(E/L_j) = D_j$. Since D_j is cyclic, the field extension E/L_j is, by Kummer theory (see, e.g., [8]), a radical extension; that is, $E = L_j(\sqrt[m_j]{\alpha_j})$ for suitable $m_j \in \mathbb{N}$ and $\alpha_j \in L_j$, provided that L_j contains all the m_jth roots of unity.

To end our discussion, let us summarize what we have learned: If $f \in \mathbb{Q}[X]$ is an arbitrary polynomial of degree n with decomposition field $E = \mathbb{Q}(\zeta_1, \ldots, \zeta_n)$, the Frobenius automorphisms of the primes unramified in E give rise to intermediate fields between $\mathbb{Q}$ and E, so that E becomes a radical extension over these intermediate fields. We thus conclude that one tries to use the fine structure of the Galois group $\mathrm{Gal}(E/\mathbb{Q})$ to describe the Galois extension $E/\mathbb{Q}$ with the help of radical extensions.

It is thus plausible that the determination of the Frobenius automorphisms plays a crucial role, something that is being investigated in current

number-theoretic research with the help of what is known as *(modular) Galois representations* of the Galois group $\mathrm{Gal}(E/\mathbb{Q})$. The prominence of this contemporary research area is revealed particularly in the fact that the theory of Galois representations played a central role in the proof of Fermat's last theorem, that is, Theorem C.8, proved by Andrew Wiles. The reader interested in pursuing this topic is referred to the survey article [7].

References

[1] N. H. Abel: *Mémoire sur les équations algébriques où l'on démontre l'impossibilité de la résolution de l'équation générale du cinquième degré.* Christiania, de l'imprimerie de Groendahl, 1824. Available online at `www.abelprisen.no/c53201/binfil/download.php?tid=53608`.

[2] E. Artin: *Galois theory.* Edited and supplemented with a section on applications by A. N. Milgram. 2nd edition, with additions and revisions, 5th reprinting. Notre Dame Mathematical Lectures, No. 2, University of Notre Dame Press, South Bend, IN, 1959.

[3] J. Bewersdorff: *Galois theory for beginners. A historical perspective.* Translated from the second German 2004 edition by D. Kramer. American Mathematical Society, Providence, RI, 2006.

[4] D. A. Cox: *Galois Theory.* John Wiley & Sons, Hoboken, NJ, 2nd edition, 2012.

[5] J.-P. Escofier: *Galois theory.* Translated from the 1997 French original by L. Schneps. Springer, Berlin Heidelberg New York, 2001.

[6] D. Jörgensen: *Der Rechenmeister.* Aufbau Taschenbuch Verlag, 6. Auflage, 2004.

[7] J. Kramer: *Über den Beweis der Fermat-Vermutung I, II.* Elem. Math. **50** (1995), 12–25; **53** (1998), 45–60.

[8] S. Lang: *Algebraic number theory.* Springer, Berlin Heidelberg New York, 2nd edition, 1994.

[9] I. Stewart: *Galois theory.* Chapman and Hall/CRC, 4th edition, 2015.

VI Hamilton's Quaternions

1. Hamilton's Quaternions as a Real Vector Space

A question of an essentially academic nature is whether the field $\mathbb{C}$ of complex numbers can be enlarged to a field that, like the field $\mathbb{C}$, is a finite-dimensional real vector space. It turns out that such an enlargement is impossible. Nevertheless, there is such a larger set of numbers if one is willing to give up commutativity of multiplication. We are then led to the skew field $\mathbb{H}$ of *quaternions*, first described by William Rowan Hamilton, which we shall present in this chapter.

We begin by defining, in addition to the complex number i, two additional imaginary units j, k such that the elements $1, i, j, k$ are linearly independent over $\mathbb{R}$. This allows us to form the 4-dimensional vector space

$$\mathbb{H} := \{\alpha = \alpha_1 \cdot 1 + \alpha_2 \cdot i + \alpha_3 \cdot j + \alpha_4 \cdot k \mid \alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R}\}.$$

Definition 1.1. We call the expression

$$\alpha = \alpha_1 \cdot 1 + \alpha_2 \cdot i + \alpha_3 \cdot j + \alpha_4 \cdot k = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k$$

a *quaternion*, and $\mathbb{H}$ the set of *quaternions*.

Remark 1.2. By construction, the sum of two quaternions $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k$ and $\beta = \beta_1 + \beta_2 i + \beta_3 j + \beta_4 k$ is given by

$$\alpha + \beta := (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2)i + (\alpha_3 + \beta_3)j + (\alpha_4 + \beta_4)k.$$

This addition is clearly associative and commutative. The additive identity element is the zero element $0 := 0 + 0i + 0j + 0k$; the additive inverse of α is given by

$$-\alpha := (-\alpha_1) + (-\alpha_2)i + (-\alpha_3)j + (-\alpha_4)k = -\alpha_1 - \alpha_2 i - \alpha_3 j - \alpha_4 k.$$

Definition 1.3. The product of two quaternions $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k$ and $\beta = \beta_1 + \beta_2 i + \beta_3 j + \beta_4 k$ is defined by

$$\begin{aligned}\alpha \cdot \beta :&= (\alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3 - \alpha_4\beta_4) + (\alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_4 - \alpha_4\beta_3)i \\ &\quad + (\alpha_1\beta_3 - \alpha_2\beta_4 + \alpha_3\beta_1 + \alpha_4\beta_2)j + (\alpha_1\beta_4 + \alpha_2\beta_3 - \alpha_3\beta_2 + \alpha_4\beta_1)k.\end{aligned}$$

Remark 1.4. We observe that this product is associative, but it is not commutative. In particular we have the following multiplication table for $1, i, j, k$:

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0_6

$$
\begin{aligned}
&i^2 = j^2 = k^2 = -1,\\
&1 \cdot i = i = i \cdot 1, \quad 1 \cdot j = j = j \cdot 1, \quad 1 \cdot k = k = k \cdot 1,\\
&i \cdot j = k = -j \cdot i, \quad j \cdot k = i = -k \cdot j, \quad k \cdot i = j = -i \cdot k.
\end{aligned}
$$

The multiplicative identity element is the unit element $1 := 1 + 0i + 0j + 0k$. It is not difficult to prove the two distributive laws.

Remark 1.5. The noncommutativity of multiplication has as a consequence the existence of polynomials in $\mathbb{H}[X]$ whose number of zeros is greater than the degree of the polynomial. Indeed, the number of zeros can be infinite.

Exercise 1.6. Verify the assertion of Remark 1.5.

Exercise 1.7. The *center* $Z(\mathbb{H})$ of $\mathbb{H}$ is defined by

$$Z(\mathbb{H}) := \{\alpha \in \mathbb{H} \mid \alpha \cdot \beta = \beta \cdot \alpha, \forall \beta \in \mathbb{H}\}.$$

Prove the equality $Z(\mathbb{H}) = \mathbb{R}$.

Remark 1.8. It is impossible to construct a 3-dimensional real vector space that contains $\mathbb{C}$ and extends the multiplication of $\mathbb{C}$. For if we choose, in addition to i, an additional imaginary number j such that the elements $1, i, j$ are linearly independent over $\mathbb{R}$, and with them, form the 3-dimensional real vector space

$$\mathbb{H}^* := \{\alpha = \alpha_1 \cdot 1 + \alpha_2 \cdot i + \alpha_3 \cdot j \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\},$$

then we must have $i \cdot j \in \mathbb{H}^*$, that is, $i \cdot j = \beta_1 \cdot 1 + \beta_2 \cdot i + \beta_3 \cdot j$ for certain $\beta_1, \beta_2, \beta_3 \in \mathbb{R}$. But that gives us the equality

$$
\begin{aligned}
(-1) \cdot j = (i \cdot i) \cdot j = i \cdot (i \cdot j) &= \beta_1 \cdot i - \beta_2 \cdot 1 + \beta_3 \cdot (i \cdot j)\\
&= \beta_1 \cdot i - \beta_2 \cdot 1 + \beta_3 \cdot (\beta_1 \cdot 1 + \beta_2 \cdot i + \beta_3 \cdot j)\\
&= (-\beta_2 + \beta_1\beta_3) \cdot 1 + (\beta_1 + \beta_2\beta_3) \cdot i + \beta_3^2 \cdot j.
\end{aligned}
$$

Since the elements $1, i, j$ are linearly independent over $\mathbb{R}$, we obtain in particular $\beta_3^2 = -1$. But this contradicts $\beta_3 \in \mathbb{R}$.

Remark 1.9. If we write a quaternion $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{H}$ in the form $\alpha = z + wj$ with $z := \alpha_1 + \alpha_2 i$, $w := \alpha_3 + \alpha_4 i \in \mathbb{C}$, then $\mathbb{H}$ can be considered a 2-dimensional complex vector space.

Definition 1.10. Let $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{H}$. The real number α_1 is called the *real part* of α and is denoted by $\mathrm{Re}(\alpha)$. The ordered triple $(\alpha_2, \alpha_3, \alpha_4)$ of real numbers is called the *imaginary part* of α and is denoted by $\mathrm{Im}(\alpha)$. If we have $\mathrm{Re}(\alpha) = 0$ and $\alpha \neq 0$, then α is said to by *purely imaginary*.

Definition 1.11. The set

$$\mathrm{Im}(\mathbb{H}) := \{\alpha_2 \cdot i + \alpha_3 \cdot j + \alpha_4 \cdot k \mid \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R}\} \subseteq \mathbb{H}$$

of all quaternions with zero real part is called the *imaginary space* of $\mathbb{H}$.

Remark 1.12. The imaginary space $\mathrm{Im}(\mathbb{H})$ is a 3-dimensional real vector space that can be identified with $\mathbb{R}^3$ via the bijective $\mathbb{R}$-linear mapping $h : \mathrm{Im}(\mathbb{H}) \longrightarrow \mathbb{R}^3$ given by

$$\alpha = \alpha_2 i + \alpha_3 j + \alpha_4 k \mapsto \mathrm{Im}(\alpha)^t = \begin{pmatrix} \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{pmatrix}.$$

Remark 1.13. For $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{H}$, we shall frequently write

$$\alpha = \mathrm{Re}(\alpha) + \mathrm{Im}(\alpha) \cdot \mathbf{i},$$

where we set $\mathbf{i} := (i, j, k)^t$.

Exercise 1.14.
(a) Show that $\mathrm{Im}(\mathbb{H}) = \{\alpha \in \mathbb{H} \mid \alpha \notin \mathbb{R} \setminus \{0\} \text{ and } \alpha^2 \in \mathbb{R}\}$.
(b) Show that $\alpha \cdot \beta + \beta \cdot \alpha \in \mathbb{R}$ for all $\alpha, \beta \in \mathrm{Im}(\mathbb{H})$.

Exercise 1.15. Prove the following formula for the product of two purely imaginary quaternions $\alpha = \mathrm{Im}(\alpha) \cdot \mathbf{i}$ and $\beta = \mathrm{Im}(\beta) \cdot \mathbf{i} \in \mathrm{Im}(\mathbb{H})$:

$$\alpha \cdot \beta = -\langle \mathrm{Im}(\alpha)^t, \mathrm{Im}(\beta)^t \rangle + (\mathrm{Im}(\alpha)^t \times \mathrm{Im}(\beta)^t) \cdot \mathbf{i},$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product on $\mathbb{R}^3$, and $\times$ is the vector product on $\mathbb{R}^3$.

Definition 1.16. In analogy to complex conjugation, we define the *conjugate quaternion* $\overline{\alpha}$ to $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{H}$ by

$$\overline{\alpha} := \alpha_1 - \alpha_2 i - \alpha_3 j - \alpha_4 k.$$

Definition 1.17. The Euclidean scalar product $\langle \cdot, \cdot \rangle : \mathbb{H} \times \mathbb{H} \longrightarrow \mathbb{R}$ is defined by

$$\langle \alpha, \beta \rangle := \mathrm{Re}(\alpha \cdot \overline{\beta}) = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 + \alpha_4\beta_4,$$

where $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k$, $\beta = \beta_1 + \beta_2 i + \beta_3 j + \beta_4 k \in \mathbb{H}$. The *modulus* $|\alpha|$ of α is then defined by

$$|\alpha| := \sqrt{\alpha \cdot \overline{\alpha}} = \sqrt{\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2}.$$

Exercise 1.18. Show that the equality $\beta \cdot \alpha \cdot \beta = 2 \cdot \langle \overline{\alpha}, \beta \rangle \cdot \beta - \langle \beta, \beta \rangle \cdot \overline{\alpha}$ holds for all $\alpha, \beta \in \mathbb{H}$.

Exercise 1.19.
(a) Show that for all $\alpha, \beta \in \mathbb{H}$, we have the equality $\overline{\alpha \cdot \beta} = \overline{\beta} \cdot \overline{\alpha}$.
(b) Show that for all $\alpha, \beta \in \mathbb{H}$, we have the *product rule* $|\alpha \cdot \beta| = |\alpha| \cdot |\beta|$.
(c) Consider how one could use part (a) to prove the following statement: if each of two natural numbers can be represented as the sum of four squares of natural numbers, then the product of those two numbers can also be represented as the sum of four squares of natural numbers.

Theorem 1.20. *The structure* $(\mathbb{H}, +, \cdot)$ *is a skew field with unit element* $1 = 1 + 0i + 0j + 0k$ *that contains the fields of real and complex numbers.*

Proof. We have only to show that every nonzero quaternion α has a multiplicative inverse. This can be easily obtained, in analogy to the complex case, by

$$\alpha^{-1} = \frac{\overline{\alpha}}{|\alpha|^2}.$$

The remaining assertions are easy to prove. □

Exercise 1.21. Complete the proof of Theorem 1.20.

An immediate consequence of Theorem 1.20 is that the quaternions possess the structure of an $\mathbb{R}$-algebra.

Definition 1.22. An $\mathbb{R}$-vector space V with a multiplication operation $\cdot : V \times V \longrightarrow V$ given by the assignment $(v_1, v_2) \mapsto v_1 \cdot v_2$ is called an $\mathbb{R}$-*algebra* if the two distributive laws

$$(\lambda_1 v_1 + \lambda_2 v_2) \cdot v_3 = \lambda_1 (v_1 \cdot v_3) + \lambda_2 (v_2 \cdot v_3),$$
$$v_1 \cdot (\lambda_1 v_2 + \lambda_2 v_3) = \lambda_1 (v_1 \cdot v_2) + \lambda_2 (v_1 \cdot v_3),$$

hold for all $\lambda_1, \lambda_2 \in \mathbb{R}$ and $v_1, v_2, v_3 \in V$. If the operation $\cdot$ is associative, that is, if $(v_1 \cdot v_2) \cdot v_3 = v_1 \cdot (v_2 \cdot v_3)$ for all $v_1, v_2, v_3 \in V$, then the $\mathbb{R}$-algebra is called an *associative* $\mathbb{R}$-algebra. If the operation $\cdot$ is commutative, that is, if $v_1 \cdot v_2 = v_2 \cdot v_1$ for all $v_1, v_2 \in V$, then the $\mathbb{R}$-algebra is called a *commutative* $\mathbb{R}$-algebra. The dimension of V as an $\mathbb{R}$-vector space is called the *dimension* of the $\mathbb{R}$-algebra V.

Definition 1.23. A nontrivial $\mathbb{R}$-algebra V is called a *division algebra* if the two equations

$$v_1 \cdot x = v_2 \quad \text{and} \quad y \cdot v_1 = v_2$$

have unique solutions x and y for $v_1, v_2 \in V$, $v_1 \neq 0$, in V.

Example 1.24. (i) The field $\mathbb{R}$ is an associative and commutative division algebra of dimension 1. The field $\mathbb{C}$ is an associative and commutative division algebra of dimension 2.

(ii) The $\mathbb{R}$-vector space $\mathbb{R}^3$ together with the vector product $\times : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ given by

$$v_1 \times v_2 := (\mu_2\nu_3 - \mu_3\nu_2, \mu_3\nu_1 - \mu_1\nu_3, \mu_1\nu_2 - \mu_2\nu_1)^t$$

for $v_1 = (\mu_1, \mu_2, \mu_3)^t$, $v_2 = (\nu_1, \nu_2, \nu_3)^t \in \mathbb{R}^3$, is an $\mathbb{R}$-algebra of dimension 3 that is neither associative nor commutative. The vector product $\times$ is, however, *anticommutative*, that is, we have $v_1 \times v_2 = -v_2 \times v_1$ for all $v_1, v_2 \in \mathbb{R}^3$.

(iii) The $\mathbb{R}$-vector space $\mathrm{M}_2(\mathbb{R})$ together with matrix multiplication is an associative $\mathbb{R}$-algebra of dimension $2^2 = 4$. The $\mathbb{R}$-vector space $\mathrm{M}_2(\mathbb{C})$ together with matrix multiplication is an associative $\mathbb{R}$-algebra of dimension $2 \cdot 2^2 = 8$. The $\mathbb{R}$-algebras $\mathrm{M}_2(\mathbb{R})$ and $\mathrm{M}_2(\mathbb{C})$ are neither commutative algebras nor division algebras.

Exercise 1.25. Verify in detail the assertions of Example 1.24.

Corollary 1.26. *The quaternions* $\mathbb{H}$ *have the structure of an associative division algebra of dimension* 4.

Proof. This result is a direct consequence of Theorem 1.20. □

Exercise 1.27. An $\mathbb{R}$-linear mapping $f : V \longrightarrow W$ of $\mathbb{R}$-algebras with multiplication operations $\cdot_V$ and $\cdot_W$ is called an $\mathbb{R}$-*algebra homomorphism* if for all $v_1, v_2 \in V$, we have the equality $f(v_1 \cdot_V v_2) = f(v_1) \cdot_W f(v_2)$. An $\mathbb{R}$-vector subspace $U \subseteq V$ is called an $\mathbb{R}$-*subalgebra* of V if $u_1 \cdot_V u_2 \in U$ for all $u_1, u_2 \in U$.

Show that the mapping $f : \mathbb{C} \longrightarrow \mathrm{M}_2(\mathbb{R})$ of Lemma 2.4 of Chapter V is an injective $\mathbb{R}$-algebra homomorphism and that the image $\mathrm{im}(f) = \mathcal{C}$ is an $\mathbb{R}$-subalgebra of $\mathrm{M}_2(\mathbb{R})$.

2. Quaternions of Modulus 1 and the Special Unitary Group

In this section, we shall identify the set of all quaternions of unit modulus with the special unitary group.

We consider the noncommutative ring $(\mathrm{M}_2(\mathbb{C}), +, \cdot)$, that is, the set of all 2×2 matrices with complex entries,

$$\mathrm{M}_2(\mathbb{C}) := \left\{ A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \,\middle|\, \alpha, \beta, \gamma, \delta \in \mathbb{C} \right\},$$

together with the usual matrix addition and multiplication. If

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{M}_2(\mathbb{C}),$$

then the conjugate matrix to A, denoted by $\overline{A}$, is defined as

$$\overline{A} := \begin{pmatrix} \overline{\alpha} & \overline{\beta} \\ \overline{\gamma} & \overline{\delta} \end{pmatrix} \in \mathrm{M}_2(\mathbb{C}).$$

Definition 2.1. We denote by

$$\mathbb{S}^3 := \{\alpha \in \mathbb{H} \mid |\alpha| = 1\}$$

the set of all quaternions of modulus 1.

Remark 2.2. The set $\mathbb{S}^3$ is a subgroup of the group $(\mathbb{H} \setminus \{0\}, \cdot)$.

Exercise 2.3. Verify the assertion of Remark 2.2.

In what follows, we shall begin by identifying the skew field of quaternions with a subring of the noncommutative ring $(\mathrm{M}_2(\mathbb{C}), +, \cdot)$.

Lemma 2.4. *The mapping* $f : (\mathbb{H}, +, \cdot) \longrightarrow (\mathrm{M}_2(\mathbb{C}), +, \cdot)$, *given by*

$$\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \mapsto \begin{pmatrix} \alpha_1 + \alpha_2 i & \alpha_3 + \alpha_4 i \\ -\alpha_3 + \alpha_4 i & \alpha_1 - \alpha_2 i \end{pmatrix},$$

is an injective ring homomorphism. The image

$$\mathcal{H} := \mathrm{im}(f) = \left\{ \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix} \,\middle|\, z, w \in \mathbb{C} \right\}$$

is a skew field. In particular, f *induces the isomorphism* $\mathbb{H} \cong \mathcal{H}$.

Proof. We begin by proving that f is an injective ring homomorphism. Let $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{H}$ and $\beta = \beta_1 + \beta_2 i + \beta_3 j + \beta_4 k \in \mathbb{H}$, and we write $\alpha \cdot \beta = \gamma_1 + \gamma_2 i + \gamma_3 j + \gamma_4 k \in \mathbb{H}$ with

$$\begin{aligned} \gamma_1 &:= \alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3 - \alpha_4\beta_4, & \gamma_2 &:= \alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_4 - \alpha_4\beta_3, \\ \gamma_3 &:= \alpha_1\beta_3 - \alpha_2\beta_4 + \alpha_3\beta_1 + \alpha_4\beta_2, & \gamma_4 &:= \alpha_1\beta_4 + \alpha_2\beta_3 - \alpha_3\beta_2 + \alpha_4\beta_1. \end{aligned}$$

We obtain

$$\begin{aligned} f(\alpha + \beta) &= f((\alpha_1 + \beta_1) + (\alpha_2 + \beta_2)i + (\alpha_3 + \beta_3)j + (\alpha_4 + \beta_4)k) \\ &= \begin{pmatrix} (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2)i & (\alpha_3 + \beta_3) + (\alpha_4 + \beta_4)i \\ -(\alpha_3 + \beta_3) + (\alpha_4 + \beta_4)i & (\alpha_1 + \beta_1) - (\alpha_2 + \beta_2)i \end{pmatrix} \\ &= \begin{pmatrix} \alpha_1 + \alpha_2 i & \alpha_3 + \alpha_4 i \\ -\alpha_3 + \alpha_4 i & \alpha_1 - \alpha_2 i \end{pmatrix} + \begin{pmatrix} \beta_1 + \beta_2 i & \beta_3 + \beta_4 i \\ -\beta_3 + \beta_4 i & \beta_1 - \beta_2 i \end{pmatrix} \\ &= f(\alpha) + f(\beta) \end{aligned}$$

and

$$\begin{aligned} f(\alpha \cdot \beta) &= f(\gamma_1 + \gamma_2 i + \gamma_3 j + \gamma_4 k) \\ &= \begin{pmatrix} \gamma_1 + \gamma_2 i & \gamma_3 + \gamma_4 i \\ -\gamma_3 + \gamma_4 i & \gamma_1 - \gamma_2 i \end{pmatrix} \\ &= \begin{pmatrix} \alpha_1 + \alpha_2 i & \alpha_3 + \alpha_4 i \\ -\alpha_3 + \alpha_4 i & \alpha_1 - \alpha_2 i \end{pmatrix} \cdot \begin{pmatrix} \beta_1 + \beta_2 i & \beta_3 + \beta_4 i \\ -\beta_3 + \beta_4 i & \beta_1 - \beta_2 i \end{pmatrix} \\ &= f(\alpha) \cdot f(\beta), \end{aligned}$$

where in the third step, we have used the equalities

$$\begin{aligned} \gamma_1 + \gamma_2 i &= (\alpha_1 + \alpha_2 i)(\beta_1 + \beta_2 i) - (\alpha_3 + \alpha_4 i)(\overline{\beta_3 + \beta_4 i}) = \overline{\gamma_1 - \gamma_2 i}, \\ \gamma_3 + \gamma_4 i &= (\alpha_1 + \alpha_2 i)(\beta_3 + \beta_4 i) + (\alpha_3 + \alpha_4 i)(\overline{\beta_1 + \beta_2 i}) = -\overline{\gamma_3 + \gamma_4 i}. \end{aligned}$$

Since we have $\ker(f) = \{0\}$, we have shown that f is an injective ring homomorphism. The image of f is given by the set

$$\mathcal{H} = \operatorname{im}(f) = \left\{ \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix} \,\middle|\, z, w \in \mathbb{C} \right\}.$$

By Lemma 3.4 of Chapter III, $\mathcal{H}$ is in fact a subring of $(\mathrm{M}_2(\mathbb{C}), +, \cdot)$. Finally, by the homomorphism theorem for rings, we have the isomorphism

$$\mathbb{H} = \mathbb{H}/\ker(f) \cong \operatorname{im}(f) = \mathcal{H}.$$

But since $\mathbb{H}$ is a skew field, it follows that $\mathcal{H}$ is also a skew field, and the proof of the lemma is complete. □

Exercise 2.5. Find additional subrings of $(\mathrm{M}_2(\mathbb{C}), +, \cdot)$ that are isomorphic to $\mathbb{H}$.

Exercise 2.6. Show that the mapping $f : \mathbb{H} \longrightarrow \mathrm{M}_2(\mathbb{C})$ of Lemma 2.4 is an injective $\mathbb{R}$-algebra homomorphism and that the image $\operatorname{im}(f) = \mathcal{H}$ is an $\mathbb{R}$-subalgebra of $\mathrm{M}_2(\mathbb{C})$ (see Exercise 1.27).

Definition 2.7. The *unitary group* $\mathrm{U}_2(\mathbb{C})$ is defined by

$$\mathrm{U}_2(\mathbb{C}) := \{A \in \mathrm{M}_2(\mathbb{C}) \mid A \cdot \overline{A}^t = E\}.$$

The *special unitary group* $\mathrm{SU}_2(\mathbb{C})$ is defined by

$$\mathrm{SU}_2(\mathbb{C}) := \{A \in \mathrm{U}_2(\mathbb{C}) \mid \det(A) = 1\}.$$

Remark 2.8. The unitary group $(\mathrm{U}_2(\mathbb{C}), \cdot)$ is a group under the operation of matrix multiplication. The special unitary group $\mathrm{SU}_2(\mathbb{C})$ is a subgroup, indeed a normal subgroup, of $(\mathrm{U}_2(\mathbb{C}), \cdot)$.

Exercise 2.9. Show that $|\det(A)| = 1$ for $A \in \mathrm{U}_2(\mathbb{C})$ and verify the assertions of Remark 2.8.

Theorem 2.10. *We have the group isomorphism*

$$(\mathbb{S}^3, \cdot) \cong (\mathrm{SU}_2(\mathbb{C}), \cdot).$$

Proof. We begin by noting that for an arbitrary matrix

$$A = \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix} \in \mathcal{H},$$

we have the equalities $\det(A) = |z|^2 + |w|^2$ and

$$A \cdot \overline{A}^t = \begin{pmatrix} z & w \\ -\overline{w} & \overline{z} \end{pmatrix} \cdot \begin{pmatrix} \overline{z} & -w \\ \overline{w} & z \end{pmatrix} = \begin{pmatrix} |z|^2 + |w|^2 & 0 \\ 0 & |z|^2 + |w|^2 \end{pmatrix} = \det(A) \cdot E.$$

If we have $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{S}^3$, that is, if $|\alpha| = 1$, then we will also have $|\alpha|^2 = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 = 1$, which yields, under the mapping f of Lemma 2.4, the equalities $\det(A) = (\alpha_1^2 + \alpha_2^2) + (\alpha_3^2 + \alpha_4^2) = 1$ and $A \cdot \overline{A}^t = \det(A) \cdot E = E$ for

$$A = f(\alpha) = \begin{pmatrix} \alpha_1 + \alpha_2 i & \alpha_3 + \alpha_4 i \\ -\alpha_3 + \alpha_4 i & \alpha_1 - \alpha_2 i \end{pmatrix}.$$

This proves that we have $A \in \mathrm{SU}_2(\mathbb{C})$. Thus the mapping f induces an injective group homomorphism $g := f|_{\mathbb{S}^3} : \mathbb{S}^3 \longrightarrow \mathrm{SU}_2(\mathbb{C})$ with image

$$\mathrm{im}(g) = \left\{ A \in \mathcal{H} \,\middle|\, \det(A) = 1 \right\} \subseteq \mathrm{SU}_2(\mathbb{C}).$$

To prove the surjectivity of g, we must show that $\mathrm{SU}_2(\mathbb{C}) \subseteq \mathrm{im}(g)$. Let

$$B := \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SO}_2(\mathbb{C}).$$

Then $B \cdot \overline{B}^t = E$, and hence $B^{-1} = \overline{B}^t$. Since $\det(B) = 1$, we have also

$$B^{-1} = \begin{pmatrix} \delta & -\beta \\ -\gamma & \alpha \end{pmatrix}.$$

We must therefore have $\delta = \overline{\alpha}$ and $\gamma = -\overline{\beta}$, which proves that we have $B \in \mathcal{H}$. Since $\det(B) = 1$, it then follows that

$$\mathrm{SU}_2(\mathbb{C}) \subseteq \left\{ A \in \mathcal{H} \,\middle|\, \det(A) = 1 \right\} = \mathrm{im}(g).$$

This completes the proof of the theorem. □

3. Quaternions of Modulus 1 and the Special Orthogonal Group

In this section, we shall identify the set of all quaternions of unit modulus with the special orthogonal group.

We show first that every quaternion of modulus 1 induces a mapping of the imaginary space into itself.

Lemma 3.1. *Let $\alpha \in \mathbb{S}^3$. The mapping $g_\alpha : \mathrm{Im}(\mathbb{H}) \longrightarrow \mathrm{Im}(\mathbb{H})$ defined by the assignment*

$$\gamma \mapsto \alpha \cdot \gamma \cdot \overline{\alpha}$$

is bijective and $\mathbb{R}$-linear. Furthermore, we have $g_{\alpha\cdot\beta} = g_\alpha \circ g_\beta$ for all $\alpha, \beta \in \mathbb{S}^3$. Moreover, the equality $g_\alpha = \mathrm{id}$ holds if and only if $\alpha \in \{\pm 1\}$.

Proof. We begin with a proof that the mapping g_α is well defined. For this, it suffices to prove the equality $\overline{g_\alpha(\gamma)} = -g_\alpha(\gamma)$ for all $\alpha \in \mathbb{S}^3$ and $\gamma \in \mathrm{Im}(\mathbb{H})$, which we obtain as follows:

$$\overline{g_\alpha(\gamma)} = \overline{\alpha \cdot \gamma \cdot \overline{\alpha}} = \overline{\gamma \cdot \overline{\alpha}} \cdot \overline{\alpha} = \alpha \cdot \overline{\gamma} \cdot \overline{\alpha} = \alpha \cdot (-\gamma) \cdot \overline{\alpha} = -g_\alpha(\gamma).$$

For all $\gamma, \delta \in \mathrm{Im}(\mathbb{H})$ and $\lambda_1, \lambda_2 \in \mathbb{R}$, we now have

$$\begin{aligned} g_\alpha(\lambda_1 \cdot \gamma + \lambda_2 \cdot \delta) &= \alpha \cdot (\lambda_1 \cdot \gamma + \lambda_2 \cdot \delta) \cdot \overline{\alpha} \\ &= \lambda_1 \cdot \alpha \cdot \gamma \cdot \overline{\alpha} + \lambda_2 \cdot \alpha \cdot \delta \cdot \overline{\alpha} \\ &= \lambda_1 \cdot g_\alpha(\gamma) + \lambda_2 \cdot g_\alpha(\delta); \end{aligned}$$

that is, g_α is $\mathbb{R}$-linear. The bijectivity of g_α is immediate from the definition of g_α.

To prove the second assertion, we note that we have the equality

$$\begin{aligned} g_{\alpha\cdot\beta}(\gamma) &= (\alpha \cdot \beta) \cdot \gamma \cdot \overline{\alpha \cdot \beta} = \alpha \cdot (\beta \cdot \gamma \cdot \overline{\beta}) \cdot \overline{\alpha} \\ &= \alpha \cdot g_\beta(\gamma) \cdot \overline{\alpha} = g_\alpha(g_\beta(\gamma)) = (g_\alpha \circ g_\beta)(\gamma) \end{aligned}$$

for all $\alpha, \beta \in \mathbb{S}^3$ and $\gamma \in \mathrm{Im}(\mathbb{H})$; that is, we have $g_{\alpha\cdot\beta} = g_\alpha \circ g_\beta$.

Finally, we obtain

$$\begin{aligned} g_\alpha(\gamma) = \gamma \quad &\Longleftrightarrow \quad \alpha \cdot \gamma \cdot \overline{\alpha} = \gamma \quad \Longleftrightarrow \quad \alpha \cdot \gamma = \gamma \cdot \alpha \\ &\Longleftrightarrow \quad \alpha \in \mathbb{R} \quad \Longleftrightarrow \quad \alpha \in \{\pm 1\} \end{aligned}$$

for all $\alpha \in \mathbb{S}^3$ and $\gamma \in \mathrm{Im}(\mathbb{H})$, where for the third equivalence, we have used the fact that $Z(\mathbb{H}) = \{\alpha \in \mathbb{H} \mid \alpha \cdot \beta = \beta \cdot \alpha, \forall \beta \in \mathbb{H}\} = \mathbb{R}$ (see Exercise 1.7). This proves the third assertion of the lemma. □

We now consider the noncommutative ring $(\mathrm{M}_3(\mathbb{R}), +, \cdot)$, that is, the set of all 3×3 matrices with real entries under the operations of matrix addition and multiplication (cf. Chapter V). We denote the unit element of $\mathrm{M}_3(\mathbb{R})$ by

$$E := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Definition 3.2. The *orthogonal group* $\mathrm{O}_3(\mathbb{R})$ is defined by

$$\mathrm{O}_3(\mathbb{R}) := \{A \in \mathrm{M}_3(\mathbb{R}) \mid A \cdot A^t = E\}.$$

The *special orthogonal group* $\mathrm{SO}_3(\mathbb{R})$ is defined by

$$\mathrm{SO}_3(\mathbb{R}) := \{A \in \mathrm{O}_3(\mathbb{R}) \mid \det(A) = 1\}.$$

Remark 3.3. The orthogonal group $(\mathrm{O}_3(\mathbb{R}), \cdot)$ is a group under matrix multiplication. The special orthogonal group $\mathrm{SO}_3(\mathbb{R})$ is a subgroup, indeed a normal subgroup, of $(\mathrm{O}_3(\mathbb{R}), \cdot)$.

Exercise 3.4. Show that $\det(A) = \pm 1$ for $A \in \mathrm{O}_3(\mathbb{R})$ and verify the assertions of Remark 3.3.

Remark 3.5. In linear algebra, one proves that every orientation-preserving rotation of $\mathbb{R}^3$ about a line passing through the origin is given by an $\mathbb{R}$-linear mapping of the form $v \mapsto A \cdot v$ ($v \in \mathbb{R}^3$) for some $A \in \mathrm{SO}_3(\mathbb{R})$. Conversely, every $\mathbb{R}$-linear mapping $v \mapsto A \cdot v$ ($v \in \mathbb{R}^3$) with $A \in \mathrm{SO}_3(\mathbb{R})$ is an orientation-preserving rotation of $\mathbb{R}^3$ about a line through the origin. With the choice of a suitable basis for $\mathbb{R}^3$, the matrix A assumes the form

$$A = E + \sin(\varphi) \cdot N + (1 - \cos(\varphi)) \cdot N^2,$$

where we have set

$$N := \begin{pmatrix} 0 & -\nu_3 & \nu_2 \\ \nu_3 & 0 & -\nu_1 \\ -\nu_2 & \nu_1 & 0 \end{pmatrix}.$$

Here $(\nu_1, \nu_2, \nu_3)^t \in \mathbb{R}^3$ is a unit vector that determines the axis of rotation, and $\varphi \in [0, 2\pi)$ is the rotation angle.

Exercise 3.6. Prove the assertions of Remark 3.5.

Theorem 3.7. *The mapping* $f : (\mathbb{S}^3, \cdot) \longrightarrow (\mathrm{SO}_3(\mathbb{R}), \cdot)$ *given by the assignment*

$$\alpha \mapsto \begin{pmatrix} \alpha_1^2 + \alpha_2^2 - \alpha_3^2 - \alpha_4^2 & 2(-\alpha_1\alpha_4 + \alpha_2\alpha_3) & 2(\alpha_1\alpha_3 + \alpha_2\alpha_4) \\ 2(\alpha_1\alpha_4 + \alpha_2\alpha_3) & \alpha_1^2 - \alpha_2^2 + \alpha_3^2 - \alpha_4^2 & 2(-\alpha_1\alpha_2 + \alpha_3\alpha_4) \\ 2(-\alpha_1\alpha_3 + \alpha_2\alpha_4) & 2(\alpha_1\alpha_2 + \alpha_3\alpha_4) & \alpha_1^2 - \alpha_2^2 - \alpha_3^2 + \alpha_4^2 \end{pmatrix},$$

where $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{S}^3$, *is a surjective group homomorphism. We*

have the group isomorphism

$$\mathbb{S}^3/\{\pm 1\} \cong \mathrm{SO}_3(\mathbb{R}).$$

Proof. We set $A_\alpha := f(\alpha)$ for $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{S}^3$. We show first that the mapping f is well defined and surjective. If $\alpha \in \{\pm 1\}$, then $A_\alpha = E \in \mathrm{SO}_3(\mathbb{R})$. If $\alpha \in \mathbb{S}^3 \setminus \{\pm 1\}$, then α has a unique representation of the form

$$\alpha = \cos\left(\frac{\varphi}{2}\right) + \sin\left(\frac{\varphi}{2}\right) \cdot \mathrm{Im}(\nu)^t \cdot \mathbf{i}$$

with $\nu = \nu_1 i + \nu_2 j + \nu_3 k \in \mathrm{Im}(\mathbb{H}) \cap \mathbb{S}^3$, where

$$\nu_j := \frac{\alpha_{j+1}}{\sqrt{\alpha_2^2 + \alpha_3^2 + \alpha_4^2}}$$

for $j = 1, 2, 3$, and $\varphi \in (0, 2\pi)$ has been chosen such that $\cos(\varphi/2) = \alpha_1$ and $\sin(\varphi/2) = (\alpha_2^2 + \alpha_3^2 + \alpha_4^2)^{1/2}$. Using the identities

$$2\alpha_1(\alpha_2^2 + \alpha_3^2 + \alpha_4^2)^{1/2} = 2\sin\left(\frac{\varphi}{2}\right)\cos\left(\frac{\varphi}{2}\right) = \sin(\varphi),$$
$$2(\alpha_2^2 + \alpha_3^2 + \alpha_4^2) = 2\sin\left(\frac{\varphi}{2}\right)^2 = 1 - \cos(\varphi),$$

we obtain the representation

$$\begin{aligned} A_\alpha &= E + 2\alpha_1 \begin{pmatrix} 0 & -\alpha_4 & \alpha_3 \\ \alpha_4 & 0 & -\alpha_2 \\ -\alpha_3 & \alpha_2 & 0 \end{pmatrix} + 2 \begin{pmatrix} -\alpha_3^2 - \alpha_4^2 & \alpha_2\alpha_3 & \alpha_2\alpha_4 \\ \alpha_2\alpha_3 & -\alpha_2^2 - \alpha_4^2 & \alpha_3\alpha_4 \\ \alpha_2\alpha_4 & \alpha_3\alpha_4 & -\alpha_2^2 - \alpha_3^2 \end{pmatrix} \\ &= E + \sin(\varphi) \cdot N_\alpha + (1 - \cos(\varphi)) \cdot N_\alpha^2, \end{aligned}$$

where we have set

$$N_\alpha := \begin{pmatrix} 0 & -\nu_3 & \nu_2 \\ \nu_3 & 0 & -\nu_1 \\ -\nu_2 & \nu_1 & 0 \end{pmatrix}.$$

Taking into account Remark 3.5, we see that this proves that $A \in \mathrm{SO}_3(\mathbb{R})$ and that f is surjective.

We shall prove the remaining assertions with the help of Lemma 3.1. To this end, let us have $\alpha \in \mathbb{S}^3$ and let $g_\alpha : \mathrm{Im}(\mathbb{H}) \longrightarrow \mathrm{Im}(\mathbb{H})$, $\gamma \mapsto \alpha \cdot \gamma \cdot \overline{\alpha}$, be the bijective $\mathbb{R}$-linear mapping from Lemma 3.1. Furthermore, let $h : \mathrm{Im}(\mathbb{H}) \longrightarrow \mathbb{R}^3$, $\gamma \mapsto \mathrm{Im}(\gamma)^t$, be the bijective $\mathbb{R}$-linear mapping from Remark 1.12. We begin by showing that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{Im}(\mathbb{H}) & \xrightarrow{h} & \mathbb{R}^3 \\ \Big\downarrow{\scriptstyle g_\alpha} & & \Big\downarrow{\scriptstyle A_\alpha} \\ \mathrm{Im}(\mathbb{H}) & \xrightarrow{h} & \mathbb{R}^3 \end{array}$$

Recall that the diagram being commutative means that $h(g_\alpha(\gamma)) = A_\alpha \cdot h(\gamma)$ for all $\gamma \in \mathrm{Im}(\mathbb{H})$. By the $\mathbb{R}$-linearity of the mappings under consideration, it suffices to prove this equality for $\gamma = i, j, k \in \mathrm{Im}(\mathbb{H})$. One can prove each of these by a short computation, which we shall demonstrate for $\gamma = i$. From

$$\begin{aligned} \alpha \cdot i \cdot \overline{\alpha} &= (-\alpha_2 + \alpha_1 i + \alpha_4 j - \alpha_3 k) \cdot (\alpha_1 - \alpha_2 i - \alpha_3 j - \alpha_4 k) = \\ &= (\alpha_1^2 + \alpha_2^2 - \alpha_3^2 - \alpha_4^2) i + 2(\alpha_1\alpha_4 + \alpha_2\alpha_3) j + 2(-\alpha_1\alpha_3 + \alpha_2\alpha_4) k \end{aligned}$$

follows, as asserted, the equality

$$\begin{aligned} h(g_\alpha(i)) &= \left(\alpha_1^2 + \alpha_2^2 - \alpha_3^2 - \alpha_4^2, 2(\alpha_1\alpha_4 + \alpha_2\alpha_3), 2(-\alpha_1\alpha_3 + \alpha_2\alpha_4)\right)^t \\ &= A_\alpha \cdot (1,0,0)^t = A_\alpha \cdot h(i). \end{aligned}$$

In the same way, we can see that $h(g_\alpha(j))$ and $h(g_\alpha(k))$ give us the second and third columns of the matrix A_α. Using Lemma 3.1, we obtain the equality

$$\begin{aligned} A_{\alpha\cdot\beta} \cdot v &= (h \circ g_{\alpha\cdot\beta} \circ h^{-1})(v) = (h \circ g_\alpha \circ g_\beta \circ h^{-1})(v) \\ &= ((h \circ g_\alpha \circ h^{-1}) \circ (h \circ g_\beta \circ h^{-1}))(v) = (A_\alpha \cdot A_\beta) \cdot v \end{aligned}$$

for all $\alpha, \beta \in \mathbb{S}^3$ and $v \in \mathbb{R}^3$; that is, we have $f(\alpha \cdot \beta) = A_{\alpha\cdot\beta} = A_\alpha \cdot A_\beta = f(\alpha) \cdot f(\beta)$. This proves that f is a group homomorphism. Moreover, from Lemma 3.1, we obtain the equivalence

$$A_\alpha \cdot v = v, \forall v \in \mathbb{R}^3 \Longleftrightarrow g_\alpha(h^{-1}(v)) = h^{-1}(v), \forall v \in \mathbb{R}^3 \Longleftrightarrow \alpha \in \{\pm 1\}.$$

This proves that $\ker(f) = \{\alpha \in \mathbb{S}^3 \mid A_\alpha = E\} = \{\pm 1\}$.

Finally, by the homomorphism theorem for groups, we have the isomorphism

$$\mathbb{S}^3/\{\pm 1\} \cong \mathrm{im}(f) = \mathrm{SO}_3(\mathbb{R}).$$

This completes the proof. □

Remark 3.8. Since we have the equality $N_\alpha \cdot v = \mathrm{Im}(\nu)^t \times v$ for all $v \in \mathbb{R}^3$, the proof of Theorem 3.7 shows that the mapping $g_\alpha : \mathrm{Im}(\mathbb{H}) \longrightarrow \mathrm{Im}(\mathbb{H})$ of Lemma 3.1 can be described by the assignment

$$\gamma \mapsto \Big(\cos(\varphi)\cdot \mathrm{Im}(\gamma)^t + \sin(\varphi)\cdot\big(\mathrm{Im}(\nu)^t \times \mathrm{Im}(\gamma)^t\big) + \big(1-\cos(\varphi)\big)\langle \mathrm{Im}(\nu)^t, \mathrm{Im}(\gamma)^t\rangle \cdot \mathrm{Im}(\nu)^t\Big)\cdot \mathbf{i}.$$

F. Extensions of Number Systems: What Comes after the Quaternions?

Beginning with the natural numbers, we have in this book systematically gone on to construct the integers, and from them, the field of rational numbers. Using the rational numbers, we then obtained the field of real numbers, which we then extended to the field of complex numbers. We have seen in this chapter that the field of complex numbers cannot be enlarged unless we abandon commutativity of multiplication. We therefore arrived at the skew field of Hamilton's quaternions, which contains the fields of real and complex numbers and possesses the structure of a real associative division algebra of dimension 4.

To conclude this book, we shall investigate in this section the question whether and to what extent this process can be continued. We shall see in particular that by further doing without associativity of multiplication, we are led to the number system of Cayley's octonions, which contains the systems $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$.

F.1 Cayley's Octonions

In October 1843, William Rowan Hamilton announced his discovery of the quaternions. In December of that same year, John Graves gave the first description of the octonions, but his work remained unpublished until 1848. They were discovered independently by Arthur Cayley, who published his results in an 1845 article, and it is on that basis that his name has become associated with the octonions. The octonions are also known as octaves (Graves's term) or Cayley numbers. For further information on the historical background of the octonions, the reader is referred, for example, to the book [8].

In order to define the octonions, we require seven imaginary units $i_1,\dots,i_7$ such that the elements $1, i_1, i_2, i_3, i_4, i_5, i_6, i_7$ are linearly independent over $\mathbb{R}$. To simplify the notation, we frequently write i_0 instead of 1.

Definition F.1. We define an 8-dimensional vector space

$$\mathbb{O} := \{\alpha = \alpha_0\cdot 1 + \alpha_1\cdot i_1 + \cdots + \alpha_7\cdot i_7 \mid \alpha_0,\dots,\alpha_7 \in \mathbb{R}\}.$$

We call $\alpha \in \mathbb{O}$ an *octonion* and $\mathbb{O}$ the set of *octonions*. The real number α_0 is called the *real part* of α and is denoted by $\mathrm{Re}(\alpha)$. The ordered 7-tuple $(\alpha_1,\dots,\alpha_7)$ of real numbers is called the *imaginary part* of α and is denoted

by $\mathrm{Im}(\alpha)$. The set

$$\mathrm{Im}(\mathbb{O}) := \{\alpha_1 \cdot i_1 + \cdots + \alpha_7 \cdot i_7 \mid \alpha_1, \ldots, \alpha_7 \in \mathbb{R}\} \subseteq \mathbb{O}$$

of all octonions with zero real part is called the *imaginary space* of $\mathbb{O}$.

Remark F.2. By construction, the sum of two octonions $\alpha = \alpha_0 \cdot 1 + \alpha_1 \cdot i_1 + \cdots + \alpha_7 \cdot i_7$ and $\beta = \beta_0 \cdot 1 + \beta_1 \cdot i_1 + \cdots + \beta_7 \cdot i_7$ is given by

$$\alpha + \beta := (\alpha_0 + \beta_0) \cdot 1 + (\alpha_1 + \beta_1) \cdot i_1 + \cdots + (\alpha_7 + \beta_7) \cdot i_7.$$

This addition is clearly associative and commutative. The additive identity element is the zero element $0 := 0 \cdot 1 + 0 \cdot i_1 + \cdots + 0 \cdot i_7$; the additive inverse of α is given by $-\alpha := -\alpha_0 \cdot 1 - \alpha_1 \cdot i_1 - \cdots - \alpha_7 \cdot i_7$.

Definition F.3. The product of two octonions $\alpha = \alpha_0 \cdot 1 + \alpha_1 \cdot i_1 + \cdots + \alpha_7 \cdot i_7$ and $\beta = \beta_0 \cdot 1 + \beta_1 \cdot i_1 + \cdots + \beta_7 \cdot i_7$ is defined by

$$\alpha \cdot \beta := \sum_{l=0}^{7} \sum_{m=0}^{7} \alpha_l \beta_m \cdot i_l i_m,$$

where the product of the elements $1, i_1, i_2, i_3, i_4, i_5, i_6, i_7$ is defined by the following table:

$\cdot$	1	i_1	i_2	i_3	i_4	i_5	i_6	i_7
1	1	i_1	i_2	i_3	i_4	i_5	i_6	i_7
i_1	i_1	-1	i_3	$-i_2$	i_5	$-i_4$	$-i_7$	i_6
i_2	i_2	$-i_3$	-1	i_1	i_6	i_7	$-i_4$	$-i_5$
i_3	i_3	i_2	$-i_1$	-1	i_7	$-i_6$	i_5	$-i_4$
i_4	i_4	$-i_5$	$-i_6$	$-i_7$	-1	i_1	i_2	i_3
i_5	i_5	i_4	$-i_7$	i_6	$-i_1$	-1	$-i_3$	i_2
i_6	i_6	i_7	i_4	$-i_5$	$-i_2$	i_3	-1	$-i_1$
i_7	i_7	$-i_6$	i_5	i_4	$-i_3$	$-i_2$	i_1	-1

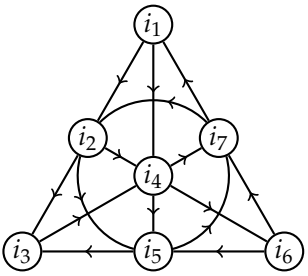
Remark F.4. The multiplicative identity element is the unit element $1 := 1 + 0 \cdot i_1 + \cdots + 0 \cdot i_7$. In addition, we have for $l, m = 1, \ldots, 7$ the equalities

$$\begin{aligned} i_l^2 &= -1, \\ 1 \cdot i_l &= i_l = i_l \cdot 1, \\ i_l \cdot i_m &= -i_m \cdot i_l \quad (l \neq m). \end{aligned}$$

This multiplication is clearly not commutative. Moreover, the multiplication is not even associative, since we have, for example, $i_4 \cdot (i_5 \cdot i_6) = -i_4 \cdot i_3 = i_7 \neq -i_7 = i_1 \cdot i_6 = (i_4 \cdot i_5) \cdot i_6$.

Remark F.5. The validity of the two distributive laws can be easily checked. The Cayley octonions $\mathbb{O}$ possess the structure of an 8-dimensional $\mathbb{R}$-algebra that is neither associative nor commutative.

Remark F.6. Multiplication of octonions can be described with the help of the Fano plane, which consists of seven points and seven lines, with each line oriented so as to form cycles, as pictured in the following diagram:



If three points i_l, i_m, i_n are located in ordered succession along a line, then one has the equality $i_l \cdot i_m = i_n$. For example, i_2, i_4, i_6 appear in order on a line, and one therefore has the equalities $i_2 \cdot i_4 = i_6$, $i_4 \cdot i_6 = i_2$, and $i_6 \cdot i_2 = i_4$. Conversely, if i_l, i_m, i_n appear in reverse order on a line, then one has the equality $i_l \cdot i_m = -i_n$. For example, i_6, i_4, i_2 appear in reverse order on a line, and one obtains the three equalities $i_6 \cdot i_4 = -i_2$, $i_4 \cdot i_2 = -i_6$, and $i_2 \cdot i_6 = -i_4$.

To see whether $\mathbb{O}$, like $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$, is a division algebra, we introduce the following definitions.

Definition F.7. Given an octonion $\alpha = \alpha_0 \cdot 1 + \alpha_1 \cdot i_1 + \cdots + \alpha_7 \cdot i_7 \in \mathbb{O}$, the *conjugate octonion* $\overline{\alpha}$ is defined by

$$\overline{\alpha} := \alpha_0 \cdot 1 - \alpha_1 \cdot i_1 - \cdots - \alpha_7 \cdot i_7.$$

If $\alpha = \alpha_0 \cdot 1 + \alpha_1 \cdot i_1 + \cdots + \alpha_7 \cdot i_7$, $\beta = \beta_0 \cdot 1 + \beta_1 \cdot i_1 + \cdots + \beta_7 \cdot i_7 \in \mathbb{O}$, we define a Euclidean scalar product $\langle \cdot, \cdot \rangle \colon \mathbb{O} \times \mathbb{O} \longrightarrow \mathbb{R}$ by setting

$$\langle \alpha, \beta \rangle := \mathrm{Re}(\alpha \cdot \overline{\beta}) = \alpha_0\beta_0 + \alpha_1\beta_1 + \cdots + \alpha_7\beta_7.$$

The *modulus* $|\alpha|$ of α is then given by

$$|\alpha| := \sqrt{\alpha \cdot \overline{\alpha}} = \sqrt{\alpha_0^2 + \alpha_1^2 + \cdots + \alpha_7^2}.$$

Remark F.8. As in Exercise 1.19, it is easily shown that for all $\alpha, \beta \in \mathbb{O}$, one has the equalities $\overline{\alpha \cdot \beta} = \overline{\beta} \cdot \overline{\alpha}$ and $\alpha \cdot \overline{\alpha} = \overline{\alpha} \cdot \alpha$.

Remark F.9. If we identify i with i_1, j with i_2, and k with i_3, we may then interpret the quaternions $\mathbb{H}$ in a natural way as an $\mathbb{R}$-subalgebra of $\mathbb{O}$. Using the identites $i_1 i_4 = i_5$, $i_2 i_4 = i_6$, and $i_3 i_4 = i_7$, we may write an octonion $\alpha = \alpha_0 \cdot 1 + \alpha_1 \cdot i_1 + \cdots + \alpha_7 \cdot i_7$ as

$$\alpha = \mathbf{a} + \mathbf{b}\, i_4, \tag{1}$$

with $\mathbf{a} := \alpha_0 \cdot 1 + \alpha_1 \cdot i_1 + \alpha_2 \cdot i_2 + \alpha_3 \cdot i_3$, $\mathbf{b} := \alpha_4 \cdot 1 + \alpha_5 \cdot i_1 + \alpha_6 \cdot i_2 + \alpha_7 \cdot i_3 \in \mathbb{H}$. For the product of two octonions $\alpha = \mathbf{a} + \mathbf{b}\, i_4$ and $\beta = \mathbf{c} + \mathbf{d}\, i_4$, we have

$$\alpha \cdot \beta = (\mathbf{ac} - \overline{\mathbf{d}}\mathbf{b}) + (\mathbf{da} + \mathbf{b}\overline{\mathbf{c}}) i_4. \tag{2}$$

Moreover, we have $\overline{\alpha} = \overline{\mathbf{a}} - \mathbf{b}\, i_4$.

Lemma F.10. *The equality*

$$\alpha(\overline{\alpha}\beta) = (\alpha\overline{\alpha})\beta \tag{3}$$

holds for all $\alpha, \beta \in \mathbb{O}$.

Proof. As in (1), we write $\alpha = \mathbf{a} + \mathbf{b}\, i_4$ and $\beta = \mathbf{c} + \mathbf{d}\, i_4$. With $\overline{\alpha} = \overline{\mathbf{a}} - \mathbf{b}\, i_4$, we obtain, using (2), the equality

$$\overline{\alpha}\beta = (\overline{\mathbf{a}}\mathbf{c} + \overline{\mathbf{d}}\mathbf{b}) + (\mathbf{d}\overline{\mathbf{a}} - \mathbf{b}\overline{\mathbf{c}}) i_4.$$

Using (2) again and taking into account the associativity of multiplication in $\mathbb{H}$, we obtain

$$\begin{aligned} \alpha(\overline{\alpha}\beta) &= \big(\mathbf{a}(\overline{\mathbf{a}}\mathbf{c} + \overline{\mathbf{d}}\mathbf{b}) - (\mathbf{a}\overline{\mathbf{d}} - \mathbf{c}\overline{\mathbf{b}})\mathbf{b}\big) + \big((\mathbf{d}\overline{\mathbf{a}} - \mathbf{b}\overline{\mathbf{c}})\mathbf{a} + \mathbf{b}(\overline{\mathbf{c}}\mathbf{a} + \overline{\mathbf{b}}\mathbf{d})\big) i_4 \\ &= \big(|\mathbf{a}|^2 + |\mathbf{b}|^2\big)\mathbf{c} + \big((|\mathbf{a}|^2 + |\mathbf{b}|^2)\mathbf{d}\big) i_4 \\ &= (\alpha\overline{\alpha})\beta, \end{aligned}$$

as claimed. □

F.2 The Octonions as a Real Division Algebra

We now proceed to show that the octonions $\mathbb{O}$ possess the structure of a real, normed, alternative division algebra.

Definition F.11. A nontrivial $\mathbb{R}$-algebra V is said to be *normed*, or a *composition algebra*, if V is a Euclidean vector space with scalar product $\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{R}$ such that

$$|v \cdot w| = |v| \cdot |w|$$

for all $v, w \in V$, called the *product rule*.

We further define a weakened form of associativity.

Definition F.12. An $\mathbb{R}$-algebra V is said to be *alternative* if the equalities

$$v(vw) = v^2 w \quad \text{and} \quad (vw)w = vw^2$$

hold for all $v, w \in V$.

Remark F.13. If an $\mathbb{R}$-algebra V is alternative, we may compute

$$\begin{aligned} 0 &= (v(v+w))(v+w) - v(v+w)^2 \\ &= (v^2 + vw)(v+w) - v(v^2 + vw + wv + w^2) \\ &= (vw)v - v(wv). \end{aligned}$$

We therefore have the equality

$$(vw)v = v(wv)$$

for all $v, w \in V$, known as the *flexible identity*.

Example F.14. The $\mathbb{R}$-algebras $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ are normed, and since multiplication is associative, they are in particular alternative.

Theorem F.15. *The structure $(\mathbb{O}, +, \cdot)$ is a real, normed, alternative division algebra of dimension 8 with unit element $1 = 1 + 0 \cdot i_1 + \cdots + 0 \cdot i_7$ that contains the fields of real and complex numbers as well as the skew field of quaternions.*

Proof. It is clear already that $(\mathbb{O}, +, \cdot)$ is an $\mathbb{R}$-algebra of dimension 8 that contains $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$. To prove that $(\mathbb{O}, +, \cdot)$ is a division algebra, we begin by showing that every nonzero octonion α possesses a multiplicative inverse. This can be shown easily in analogy to the case of the complex numbers or quaternions by writing

$$\alpha^{-1} = \frac{\overline{\alpha}}{|\alpha|^2}.$$

By Definition 1.23, we have now to show that the two equalities

$$\alpha \cdot \xi = \beta \quad \text{and} \quad \eta \cdot \alpha = \beta,$$

for $\alpha, \beta \in \mathbb{O}$ with $\alpha \neq 0$, have unique solutions $\xi, \eta \in \mathbb{O}$. To this end, we shall use the identity (3); setting $\xi := \alpha^{-1}\beta = |\alpha|^{-2}\overline{\alpha}\beta$, we then calculate

$$\alpha \cdot \xi = |\alpha|^{-2}\alpha(\overline{\alpha}\beta) = |\alpha|^{-2}(\alpha\overline{\alpha})\beta = \beta.$$

Through conjugation of (3), we obtain the identity $(\overline{\beta}\alpha)\overline{\alpha} = \overline{\beta}(\alpha\overline{\alpha})$, which leads to the equality

$$(\beta\overline{\alpha})\alpha = \beta(\overline{\alpha}\alpha), \tag{4}$$

valid for all $\alpha, \beta \in \mathbb{O}$, since along with $\alpha, \beta \in \mathbb{O}$, $\overline{\alpha}, \overline{\beta}$ run through all elements of $\mathbb{O}$. Setting $\eta := \beta\alpha^{-1} = |\alpha|^{-2}\beta\overline{\alpha}$, we thereby obtain

$$\eta \cdot \alpha = |\alpha|^{-2}(\beta\overline{\alpha})\alpha = |\alpha|^{-2}\beta(\overline{\alpha}\alpha) = \beta.$$

To prove that $(\mathbb{O}, +, \cdot)$ is an alternative algebra, we substitute $\overline{\alpha} = 2\,\mathrm{Re}(\alpha) - \alpha$ in (3) and obtain the identity

$$2\,\mathrm{Re}(\alpha)\alpha\beta - \alpha(\alpha\beta) = 2\,\mathrm{Re}(\alpha)\alpha\beta - \alpha^2\beta,$$

which leads to the identity

$$\alpha(\alpha\beta) = \alpha^2\beta$$

for all $\alpha, \beta \in \mathbb{O}$. If we now substitute $\overline{\alpha} = 2\,\mathrm{Re}(\alpha) - \alpha$ in (4), we obtain analogously the equality

$$(\beta\alpha)\alpha = \beta\alpha^2$$

for all $\alpha, \beta \in \mathbb{O}$. We have thereby shown that $\mathbb{O}$ is alternative.

Finally, to show that the product rule holds, we first establish the identity

$$(\alpha\beta)(\overline{\beta}\overline{\alpha}) = \alpha((\beta\overline{\beta})\overline{\alpha}) \tag{5}$$

for all $\alpha, \beta \in \mathbb{O}$. Since this equality holds trivially for $\alpha = 0$, we may assume that $\alpha \neq 0$, in which case there exists the multiplicative inverse α^{-1}, and since $\mathbb{O}$ is a division algebra, it suffices to show that

$$((\alpha\beta)(\overline{\beta}\overline{\alpha}))\alpha = (\alpha((\beta\overline{\beta})\overline{\alpha}))\alpha.$$

To prove this equality, we first calculate

$$\begin{aligned}((\alpha\beta)(\overline{\beta}\overline{\alpha}))\alpha &= ((\alpha\beta)(\overline{\alpha\beta}))\alpha = (\alpha\beta)((\overline{\alpha\beta})\alpha) = (\alpha\beta)((\overline{\beta}\overline{\alpha})\alpha)\\ &= (\alpha\beta)(\overline{\beta}(\overline{\alpha}\alpha)) = |\alpha|^2(\alpha\beta)\overline{\beta} = |\alpha|^2\alpha(\beta\overline{\beta}) = |\alpha|^2|\beta|^2\alpha,\end{aligned}$$

where for the second equality, we have used (3), and for the fourth and sixth equalities, (4). But we also have

$$(\alpha((\beta\overline{\beta})\overline{\alpha}))\alpha = (|\beta|^2\alpha\overline{\alpha})\alpha = |\alpha|^2|\beta|^2\alpha.$$

Altogether, this proves the equality (5). Finally, we calculate

$$|\alpha\beta|^2 = (\alpha\beta)(\overline{\alpha\beta}) = (\alpha\beta)(\overline{\beta}\overline{\alpha}) = \alpha((\beta\overline{\beta})\overline{\alpha}) = |\beta|^2\alpha\overline{\alpha} = |\alpha|^2|\beta|^2,$$

where for the third equality, we have used (5). This completes the proof. □

Remark F.16. The product rule can be proved directly, but doing so is tedious.

Remark F.17. We note that by making suitable use of the product rule, similarly to what was done in Exercise 1.19, one may prove the following assertion: If each of two natural numbers can be represented as the sum of eight squares of natural numbers, then the product of those two numbers can also be represented as the sum of eight squares of natural numbers.

F.3 Normed $\mathbb{R}$-Algebras

In this section, let $(V, \langle\cdot,\cdot\rangle)$ be a finite-dimensional normed $\mathbb{R}$-algebra with unit element 1. Conjugation on V is defined by

$$\overline{v} := 2\langle v, 1\rangle - v \tag{6}$$

for all $v \in V$, from which one immediately deduces the relationship $\overline{\overline{v}} = v$.

The validity of the product rule in its squared form

$$|v \cdot w|^2 = |v|^2 \cdot |w|^2 \quad (v, w \in V) \tag{7}$$

has several immediate consequences, which we shall proceed to list.

Lemma F.18. *For all $t, u, v, w \in V$, the following equalities hold:*
(i) $\langle uv, uw\rangle = |u|^2\langle v, w\rangle$ *and* $\langle vu, wu\rangle = |u|^2\langle v, w\rangle$.
(ii) $\langle tv, uw\rangle = 2\langle t, u\rangle\langle v, w\rangle - \langle uv, tw\rangle$.
(iii) $\langle tv, w\rangle = \langle v, \overline{t}w\rangle$ *and* $\langle tv, w\rangle = \langle t, w\overline{v}\rangle$.
(iv) $\overline{\overline{v}} = v$ *and* $\overline{v}\overline{w} = \overline{w} \cdot \overline{v}$.

Proof. To prove (i), we consider the relationship

$$|v + w|^2 = |v|^2 + |w|^2 + 2\langle v, w\rangle \quad (v, w \in V), \tag{8}$$

from which we conclude that

$$|uv + uw|^2 = |uv|^2 + |uw|^2 + 2\langle uv, uw\rangle. \tag{9}$$

If, on the other hand, we replace v in (7) by u, and w by $v + w$, we obtain, taking into account the distributive law,

$$|uv+uw|^2 = |u|^2\cdot|v+w|^2 = |u|^2\cdot\big(|v|^2+|w|^2+2\langle v,w\rangle\big), \qquad (10)$$

where for the last equality we have again made use of (8). Comparing (9) with (10) yields

$$|uv|^2+|uw|^2+2\langle uv,uw\rangle = |u|^2|v|^2+|u|^2|w|^2+2|u|^2\langle v,w\rangle,$$

which gives us the first equality on application of the product rule (7) and division by 2. The second equality can be proved analogously.

To prove (ii), we begin by using (i) to write

$$\big(|t|^2+|u|^2\big)\langle v,w\rangle = \langle tv,tw\rangle+\langle uv,uw\rangle,$$

thereby obtaining, again considering (i), the equality

$$\begin{aligned}\big(|t|^2+|u|^2\big)\langle v,w\rangle+\langle tv,uw\rangle+\langle uv,tw\rangle &= \langle (t+u)v,(t+u)w\rangle\\ &= |t+u|^2\langle v,w\rangle = \big(|t|^2+|u|^2+2\langle t,u\rangle\big)\langle v,w\rangle,\end{aligned}$$

which implies the assertion.

To prove (iii), we set $u=1$ in (ii), which yields

$$\langle tv,w\rangle = 2\langle t,1\rangle\langle v,w\rangle-\langle v,tw\rangle = \langle v,(2\langle t,1\rangle-t)w\rangle = \langle v,\bar{t}w\rangle,$$

which proves the first equality. The second equality can be proved analogously.

To prove (iv), we have now only to prove the second equality, since the first has already been established. To this end, consider, using (iii), the equality

$$\langle\overline{vw},r\rangle = \langle\bar{r}\cdot\overline{vw},1\rangle = \langle\bar{r},vw\rangle = \langle\bar{v}\cdot\bar{r},w\rangle = \langle\bar{v},wr\rangle = \langle\overline{w}\cdot\bar{v},r\rangle$$

for arbitrary $r\in V$, from which we conclude that $\overline{vw}=\overline{w}\cdot\bar{v}$. This completes the proof of the lemma. □

Definition F.19. The *imaginary space* of V is defined as the set

$$\mathrm{Im}(V) = \{v\in V \mid v\notin\mathbb{R}\setminus\{0\} \text{ and } v^2\in\mathbb{R}\}.$$

Remark F.20. (i) We have $\mathrm{Im}(V)\cap\mathbb{R}=\{0\}$, and $v\in\mathrm{Im}(V)$ implies $\lambda v\in\mathrm{Im}(V)$ for all $\lambda\in\mathbb{R}$.

(ii) Frobenius's lemma says that $\mathrm{Im}(V)$ is an $\mathbb{R}$-vector space and that $V=\mathbb{R}\oplus\mathrm{Im}(V)$. In particular, we have the equivalence

$$v\in\mathrm{Im}(V) \quad\Longleftrightarrow\quad \langle v,1\rangle=0,$$

and therefore the equality $\bar{v}=2\langle v,1\rangle-v=-v$ for all $v\in\mathrm{Im}(V)$.

Theorem F.21. *Every finite-dimensional normed $\mathbb{R}$-algebra with unit element* 1 *is an alternative division algebra.*

Proof. First of all, by Lemma F.18 (iii) and (i), we have the respective equalities

$$\langle \overline{v}(vw), r\rangle = \langle vw, vr\rangle = |v|^2\langle w, r\rangle,$$
$$\langle (wv)\overline{v}, r\rangle = \langle wv, rv\rangle = |v|^2\langle w, r\rangle,$$

for all $r \in V$, from which we obtain

$$\overline{v}(vw) = |v|^2 w = (wv)\overline{v} \tag{11}$$

for all $v, w \in V$. For $v \in V$ with $v \neq 0$, we set $v^{-1} := \overline{v}/|v|^2$. Multiplication of (11) by $1/|v|^2$ yields the equivalent equality

$$v^{-1}(vw) = w = (wv)v^{-1}, \tag{12}$$

which on replacing v with v^{-1} and taking account of $(v^{-1})^{-1} = v$ leads to the equality

$$v(v^{-1}w) = w = (wv^{-1})v.$$

We have thereby proved that the two equations

$$v \cdot x = w \quad \text{and} \quad y \cdot v = w$$

for $v, w \in V$ with $v \neq 0$ possess the unique solutions $x := v^{-1}w \in V$ and $y := wv^{-1} \in V$, which completes the proof that V is a division algebra.

If we now substitute $\overline{v} = 2\langle v, 1\rangle - v$ in (11), we obtain $(2\langle v, 1\rangle - v)(vw) = (v(2\langle v, 1\rangle - v))w$, from which we conclude at once that

$$v(vw) = v^2 w$$

for all $v, w \in V$. We prove analogously that

$$(vw)w = vw^2$$

for all $v, w \in V$. Thus we have shown that V is an alternative $\mathbb{R}$-algebra. □

Finally, we prove a theorem that will be of great importance in the following section. We recall that an $\mathbb{R}$-subalgebra of V is an $\mathbb{R}$-vector subspace $U \subseteq V$ such that for all $u_1, u_2 \in U$, the product $u_1 \cdot u_2$ is also in U (see Exercise 1.27).

Theorem F.22. *Let V be a finite-dimensional normed $\mathbb{R}$-algebra with unit element* 1*, and $U \subsetneq V$ a proper $\mathbb{R}$-subalgebra of V with $1 \in U$. Then there exists $\mathbf{i} = \mathbf{i}_U \in V$*

with the property that

$$\mathbf{i}^2 = -1 \quad \textit{and} \quad \langle \mathbf{i}, u \rangle = 0 \tag{13}$$

for all $u \in U$. In particular, we then have $\bar{\mathbf{i}} = -\mathbf{i}$ and $|\mathbf{i}| = 1$.

Proof. We have $V = \mathbb{R} \oplus \operatorname{Im}(V)$ and thereby also $U = \mathbb{R} \oplus \operatorname{Im}(U)$, from which on account of $U \subsetneq V$, we have in particular also $\operatorname{Im}(U) \subsetneq \operatorname{Im}(V)$. There exists, therefore, an element $v_0 \in \operatorname{Im}(V)$ with $v_0 \neq 0$ that satisfies

$$\langle v_0, u \rangle = 0 \tag{14}$$

for all $u \in \operatorname{Im}(U)$. Since $v_0 \in \operatorname{Im}(V)$, there exists as well $r \in \mathbb{R}$ with $v_0^2 = r$. We assert that we must have $r < 0$, for otherwise, we would obtain the relationship

$$\left(v_0 - \sqrt{r}\right)\left(v_0 + \sqrt{r}\right) = v_0^2 - r = 0,$$

from which we conclude that $v_0 \in \mathbb{R}$, because V has no zero divisors, which contradicts that $v_0 \in \operatorname{Im}(V)$ with $v_0 \neq 0$. We have thus shown that $v_0^2 = -|r|$ for a nonzero number $r \in \mathbb{R}$. Defining

$$\mathbf{i} := \frac{1}{\sqrt{|r|}} v_0 \in \operatorname{Im}(V),$$

we see at once that $\mathbf{i}^2 = -1$. If now $u \in U = \mathbb{R} \oplus \operatorname{Im}(U)$, then we may represent u in the form $u = u_1 + u_2$ with $u_1 \in \mathbb{R}$ and $u_2 \in \operatorname{Im}(U)$. Taking into account (14), we thereby obtain

$$\langle \mathbf{i}, u \rangle = \langle \mathbf{i}, u_1 \rangle + \langle \mathbf{i}, u_2 \rangle = 0 + 0 = 0,$$

as desired. Finally, we calculate

$$\bar{\mathbf{i}} = 2\langle \mathbf{i}, 1 \rangle - \mathbf{i} = -\mathbf{i}$$

and obtain thereby, since $\mathbf{i}^2 = -1$, the equality

$$|\mathbf{i}| = \sqrt{\mathbf{i}\bar{\mathbf{i}}} = \sqrt{-\mathbf{i}^2} = 1,$$

which completes the proof of the theorem. □

F.4 Hurwitz's Theorem

In this final section we prove Hurwitz's theorem, which states that up to isomorphism, the only finite-dimensional normed $\mathbb{R}$-algebras with unit el-

ement 1 are the $\mathbb{R}$-algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$. To prove this, we shall use the so-called Cayley–Dickson doubling process.

Proposition F.23. *Let V be a finite-dimensional normed $\mathbb{R}$-algebra with unit element 1. Let $U \subsetneq V$ be a proper $\mathbb{R}$-subalgebra of V with $1 \in U$, and let $\mathbf{i} = \mathbf{i}_U \in V$ be an element with the property* (13)*. Then for all $a,b,c,d \in U$, the following conditions are satisfied:*
(i) $\langle a + b\mathbf{i}, c + d\mathbf{i}\rangle = \langle a,c\rangle + \langle b,d\rangle$.
(ii) $\overline{a + b\mathbf{i}} = \overline{a} - b\mathbf{i}$.
(iii) $b\mathbf{i} = \mathbf{i}\overline{b}$.
(iv) $(a + b\mathbf{i}) \cdot (c + d\mathbf{i}) = (ac - \overline{d}b) + (da + b\overline{c})\mathbf{i}$.
In particular, $U + U\mathbf{i}$ is an $\mathbb{R}$-subalgebra of V.

Proof. To prove (i), we verify, using Lemma F.18 (iii) and (i) as well as properties (13) of $\mathbf{i}$ the equalities

$$\langle a, d\mathbf{i}\rangle = \langle \overline{d}a, \mathbf{i}\rangle = 0, \quad \langle b\mathbf{i}, c\rangle = \langle \mathbf{i}, \overline{b}c\rangle = 0,$$

and

$$\langle b\mathbf{i}, d\mathbf{i}\rangle = |\mathbf{i}|^2\langle b,d\rangle = \langle b,d\rangle.$$

This gives us the first assertion.

For the proof of (ii), we consider, taking into account Lemma F.18 (iii) and (13), the equality

$$\overline{b\mathbf{i}} = 2\langle b\mathbf{i}, 1\rangle - b\mathbf{i} = 2\langle \mathbf{i}, \overline{b}\rangle - b\mathbf{i} = -b\mathbf{i},$$

from which $\overline{a + b\mathbf{i}} = \overline{a} + \overline{b\mathbf{i}} = \overline{a} - b\mathbf{i}$ follows at once.

To prove (iii), we take $a = 0$ in (ii) and obtain, with the additional help of Lemma F.18 (iv), the equality

$$-b\mathbf{i} = \overline{b\mathbf{i}} = \overline{\mathbf{i}} \cdot \overline{b} = -\mathbf{i}\overline{b}.$$

This gives the third assertion.

To prove (iv), we first calculate

$$(a + b\mathbf{i}) \cdot (c + d\mathbf{i}) = ac + a(d\mathbf{i}) + (b\mathbf{i})c + (b\mathbf{i})(d\mathbf{i}).$$

For an arbitrary $v \in V$ we obtain, using Lemma F.18 (iii), the equality (iii) just proved, and Lemma F.18 (ii), the equality

$$\langle a(d\mathbf{i}), v\rangle = \langle d\mathbf{i}, \overline{a}v\rangle = \langle \mathbf{i}\overline{d}, \overline{a}v\rangle = 0 - \langle \overline{a}\overline{d}, \mathbf{i}v\rangle = \langle \mathbf{i}(\overline{a}\overline{d}), v\rangle = \langle (da)\mathbf{i}, v\rangle,$$

where we note that $\langle \mathbf{i}, \overline{a}\rangle = 0$. We conclude from this that $a(d\mathbf{i}) = (da)\mathbf{i}$. We verify analogously, for an arbitrary $v \in V$, using again Lemma F.18 (iii) and Lemma F.18 (ii), the equality

$$\langle (b\mathbf{i})c, v\rangle = \langle b\mathbf{i}, v\bar{c}\rangle = 0 - \langle v\mathbf{i}, b\bar{c}\rangle = \langle v, (b\bar{c})\mathbf{i}\rangle = \langle (b\bar{c})\mathbf{i}, v\rangle,$$

where we have noted that $\langle \mathbf{i}, \bar{c}\rangle = 0$. From this we conclude that $(b\mathbf{i})c = (b\bar{c})\mathbf{i}$. Finally, we obtain analogously for an arbitrary $v \in V$, using Lemma F.18 (iii), the just proved equality (iii), and Lemma F.18 (ii), the equality

$$\langle (b\mathbf{i})(d\mathbf{i}), v\rangle = \langle b\mathbf{i}, v \cdot \overline{d\mathbf{i}}\rangle = -\langle \mathbf{i}\bar{b}, v(\mathbf{i}\bar{d})\rangle = 0 + \langle v\bar{b}, \mathbf{i}(\mathbf{i}\bar{d})\rangle,$$

where we note that $\langle \bar{b}, \mathbf{i}\bar{d}\rangle = 0$. A further application of Lemma F.18 (iii) and (i) now gives

$$\langle (b\mathbf{i})(d\mathbf{i}), v\rangle = -\langle \mathbf{i}(v\bar{b}), \mathbf{i}\bar{d}\rangle = -|\mathbf{i}|^2\langle v\bar{b}, \bar{d}\rangle = -\langle v, \bar{d}b\rangle = \langle -\bar{d}b, v\rangle,$$

from which we obtain $(b\mathbf{i})(d\mathbf{i}) = -\bar{d}b$. Altogether, we thereby obtain

$$(a + b\mathbf{i}) \cdot (c + d\mathbf{i}) = ac + (da)\mathbf{i} + (b\bar{c})\mathbf{i} - \bar{d}b,$$

as asserted. This proves in particular that $U + U\mathbf{i}$ is an $\mathbb{R}$-subalgebra of V. □

Remark F.24. The $\mathbb{R}$-subalgebra $U + U\mathbf{i}$ of V from Proposition F.23 is also called the *Cayley–Dickson double of U (with respect to $\mathbf{i}$)*. The above proposition shows that whenever an $\mathbb{R}$-algebra V possesses a proper $\mathbb{R}$-subalgebra, it must also contain its Cayley–Dickson double. If V is finite-dimensional, then V itself must arise from its smallest $\mathbb{R}$-subalgebra by a finite Cayley–Dickson doubling process. In the following theorem, we shall see that the property of being normed allows only three Cayley–Dickson doubling processes.

Theorem F.25. *Let V be a finite-dimensional normed $\mathbb{R}$-algebra with unit element* 1. *In addition, let $U \subsetneq V$ be a proper $\mathbb{R}$-subalgebra of V with $1 \in U$, and let $\mathbf{i} = \mathbf{i}_U \in V$ be an element with properties* (13). *Then the following hold for the Cayley–Dickson double $U + U\mathbf{i}$*:

(i) *$U + U\mathbf{i}$ is normed if and only if U is normed and associative.*
(ii) *$U + U\mathbf{i}$ is normed and associative if and only if U is normed, associative, and commutative.*
(iii) *$U + U\mathbf{i}$ is normed, associative, and commutative if and only if U is normed, associative, commutative, and invariant under conjugation.*

Proof. (i) If $U + U\mathbf{i}$ is normed, then U is normed. Furthermore, by Proposition F.23 (iv), the $\mathbb{R}$-algebra $U + U\mathbf{i}$ is normed if and only if

$$|a + b\mathbf{i}|^2 \cdot |c + d\mathbf{i}|^2 = |(ac - \bar{d}b) + (da + b\bar{c})\mathbf{i}|^2$$

for all $a, b, c, d \in U$. By (8), (13), and Lemma F.18 (iii), this equality is equivalent to

$$\begin{aligned}(|a|^2+|b|^2)\cdot(|c|^2+|d|^2) &= |ac-\overline{d}b|^2+|da+b\overline{c}|^2\\ \iff\quad 0 &= -\langle ac,\overline{d}b\rangle+\langle da,b\overline{c}\rangle\\ \iff\quad \langle d(ac),b\rangle &= \langle (da)c,b\rangle,\end{aligned}$$

where we have made use of the product rule in U. This proves that $U+U\mathbf{i}$ is normed if and only if $d(ac)=(da)c$ for all $a,c,d\in U$, that is, if and only if U is associative. Conversely, if U is normed and associative, then the above equivalences show that then $U+U\mathbf{i}$ is normed as well.

(ii) If $U+U\mathbf{i}$ is normed and associative, then U is also normed and associative. Moreover, by Proposition F.23 (iv), the equality

$$(ad)\mathbf{i}=a(d\mathbf{i})=(da)\mathbf{i}$$

holds for all $a,d\in U$, which means that U is commutative. Conversely, if U is normed, associative, and commutative, then by (i), it remains to show that $U+U\mathbf{i}$ is associative. To this end, we compute for $a,b,c,d,e,f\in U$ the equalities

$$\begin{aligned}\big((a+b\mathbf{i})\cdot(c+d\mathbf{i})\big)\cdot(e+f\mathbf{i}) &= \big((ac-\overline{d}b)+(da+b\overline{c})\mathbf{i}\big)\cdot(e+f\mathbf{i})\\ &= (ac-\overline{d}b)e-\overline{f}(da+b\overline{c})+\big(f(ac-\overline{d}b)+(da+b\overline{c})\overline{e}\big)\mathbf{i}\\ &= (ac)e-(\overline{d}b)e-\overline{f}(da)-\overline{f}(b\overline{c})+\big(f(ac)-f(\overline{d}b)+(da)\overline{e}+(b\overline{c})\overline{e}\big)\mathbf{i}\end{aligned}$$

and

$$\begin{aligned}(a+b\mathbf{i})\cdot\big((c+d\mathbf{i})\cdot(e+f\mathbf{i})\big) &= (a+b\mathbf{i})\cdot\big((ce-\overline{f}d)+(fc+d\overline{e})\mathbf{i}\big)\\ &= a(ce-\overline{f}d)-\overline{(fc+d\overline{e})}b+\big((fc+d\overline{e})a+b\overline{(ce-\overline{f}d)}\big)\mathbf{i}\\ &= a(ce)-a(\overline{f}d)-(\overline{c}\overline{f})b-(e\overline{d})b+\big((fc)a+(d\overline{e})a+b(\overline{e}\,\overline{c})-b(\overline{d}f)\big)\mathbf{i}.\end{aligned}$$

Since U is associative and commutative, these two expressions agree. This proves that $U+U\mathbf{i}$ is associative.

(iii) If $U+U\mathbf{i}$ is normed, associative, and commutative, then U is also normed, associative, and commutative. Moreover, we have then by Proposition F.23 (iii) the equality

$$\mathbf{i}a=a\mathbf{i}=\mathbf{i}\overline{a}$$

for all $a\in U$; that is, we have $a=\overline{a}$ for all $a\in U$. Therefore, U is invariant under conjugation. Conversely, if U is normed, associative, commutative, and invariant under conjugation, then by (ii), it remains to show that $U+U\mathbf{i}$ is commutative. To prove this, we calculate for $a,b,c,d\in U$ the equalities

$$\begin{aligned}(a+b\mathbf{i})\cdot(c+d\mathbf{i}) &= (ac-\overline{d}b)+(da+b\overline{c})\mathbf{i},\\ (c+d\mathbf{i})\cdot(a+b\mathbf{i}) &= (ca-\overline{b}d)+(bc+d\overline{a})\mathbf{i}.\end{aligned}$$

Since U is associative, commutative, and invariant under conjugation, the two expressions agree. This proves that $U + U\mathbf{i}$ is commutative. □

Example F.26. Starting with the 1-dimensional associative, commutative, and conjugation-invariant $\mathbb{R}$-algebra $\mathbb{R}$, we obtain by the Cayley–Dickson doubling process the 2-dimensional associative and commutative $\mathbb{R}$-algebra $\mathbb{C} \cong \mathbb{R} + \mathbb{R}i$. If in a further step we apply the Cayley–Dickson doubling process to the $\mathbb{R}$-algebra of complex numbers $\mathbb{C}$, we will arrive at the 4-dimensional associative $\mathbb{R}$-algebra $\mathbb{H} \cong \mathbb{C} + \mathbb{C}j$. By Remark F.9, we obtain finally in a third Cayley–Dickson step applied to the $\mathbb{R}$-algebra of quaternions the 8-dimensional algebra $\mathbb{O} \cong \mathbb{H} + \mathbb{H}i_4$ of octonions. This example demonstrates quite clearly the successive loss of conjugation-invariance, commutativity, and associativity through repeated application of the Cayley–Dickson doubling process.

Theorem F.27 (Hurwitz's theorem). *The $\mathbb{R}$-algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$ are (up to isomorphism) the only finite-dimensional normed $\mathbb{R}$-algebras with unit element* 1.

Proof. Let V be a finite-dimensional normed $\mathbb{R}$-algebra with unit element 1. We shall show that V is isomorphic to $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, or $\mathbb{O}$.

If V is 1-dimensional, then V is isomorphic to $\mathbb{R}$. Otherwise, V has a proper $\mathbb{R}$-subalgebra $U_1 \subsetneq V$ that is isomorphic to $\mathbb{R}$. We have $1 \in U_1$, and by Theorem F.22, there exists $\mathbf{i} = \mathbf{i}_{U_1} \in V$ satisfying properties (13). By Theorem F.25, the Cayley–Dickson double $U_1 + U_1\mathbf{i}_{U_1}$ is a normed, associative, and commutative $\mathbb{R}$-subalgebra of V, since $U_1 \cong \mathbb{R}$ is normed, associative, commutative, and invariant under conjugation. If we now have $V = U_1 + U_1\mathbf{i}_{U_1}$, then V is isomorphic to $\mathbb{R} + \mathbb{R}\mathbf{i}_{U_1} \cong \mathbb{C}$.

Otherwise, V possesses a proper $\mathbb{R}$-subalgebra $U_2 \subsetneq V$ that is isomorphic to $\mathbb{C}$. We again have $1 \in U_2$, and by Theorem F.22, there exists $\mathbf{i} = \mathbf{i}_{U_2} \in V$ satisfying properties (13). By Theorem F.25, the Cayley–Dickson double $U_2 + U_2\mathbf{i}_{U_2}$ is a normed and associative $\mathbb{R}$-subalgebra of V, since $U_2 \cong \mathbb{C}$ is normed, associative, and commutative. If we now have $V = U_2 + U_2\mathbf{i}_{U_2}$, then V is isomorphic to $\mathbb{C} + \mathbb{C}\mathbf{i}_{U_2} \cong \mathbb{H}$.

Otherwise, V possesses a proper $\mathbb{R}$-subalgebra $U_3 \subsetneq V$ isomorphic to $\mathbb{H}$. We again have $1 \in U_3$, and by Theorem F.22, there exists $\mathbf{i} = \mathbf{i}_{U_3} \in V$ satisfying properties (13). By Theorem F.25, the Cayley–Dickson double $U_3 + U_3\mathbf{i}_{U_3}$ is a normed $\mathbb{R}$-subalgebra of V, since $U_3 \cong \mathbb{H}$ is normed and associative. If we now have $V = U_3 + U_3\mathbf{i}_{U_3}$, then V is isomorphic to $\mathbb{H} + \mathbb{H}\mathbf{i}_{U_3} \cong \mathbb{O}$.

If now V were to contain a proper $\mathbb{R}$-subalgebra $U_4 \subsetneq V$ isomorphic to $\mathbb{O}$, then we would again have $1 \in U_4$, and by Theorem F.22, there would exist $\mathbf{i} = \mathbf{i}_{U_4} \in V$ satisfying properties (13). By Theorem F.25, the Cayley–Dickson double $U_4 + U_4\mathbf{i}_{U_4}$ would not, however, be a normed $\mathbb{R}$-subalgebra of V, since $U_4 \cong \mathbb{O}$ is not associative. This contradicts the fact that V is normed. This case can therefore not occur, and the theorem is proved. □

Remark F.28. The statement of Hurwitz's theorem holds more generally for finite-dimensional real division algebras that are not necessarily normed. In this context, Heinz Hopf showed in [5], already in 1940, that the dimension of such a division algebra must be a power of 2. In 1958, Michel Kervaire and John Milnor proved in [7], independently of each other, that every finite-dimensional real division algebra has dimension 1, 2, 4, or 8. It then follows from this more general assumption that every such division algebra is isomorphic to $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, or $\mathbb{O}$. The proof of the deep theorem of Kervaire–Milnor uses methods of algebraic topology (see Chapter 10, contributed by Friedrich Hirzebruch, in [4]). As of today, there is no known purely algebraic proof.

If one applies the Cayley–Dickson doubling process again to the skew field of octonions, one obtains the 16-dimensional $\mathbb{R}$-algebra $\mathbb{S}$ of *sedenions*, which is neither commutative nor alternative; it is also not associative, and it has zero divisors, and is therefore, as we know by now, no longer a division algebra.

Remark F.29. We end this section with the observation that there is an alternative way to construct higher-dimensional $\mathbb{R}$-algebras containing $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ beginning with the real numbers and passing to the complex numbers and the quaternions. To do so, we begin with an n-dimensional Euclidean $\mathbb{R}$-vector space $(V, \langle \cdot, \cdot \rangle)$; we note that we could as well begin more generally with a K-vector space, over an arbitrary field K, and an arbitrary scalar product. We then form the tensor algebra $T(V)$ and consider its ideal

$$I(V) = \big\langle v \otimes w + w \otimes v + 2 \cdot \langle v, w \rangle \,\big|\, v, w \in V \big\rangle.$$

We then define the so-called *Clifford algebra*

$$C(V, \langle \cdot, \cdot \rangle) := T(V)/I(V),$$

in which we denote the product by $\cdot$ as usual. The Clifford algebra is by construction an associative $\mathbb{R}$-algebra of dimension 2^n. Namely if $\{e_1, \ldots, e_n\}$ is an ordered orthonormal basis of V, then $C(V, \langle \cdot, \cdot \rangle)$ has the basis

$$\{1; e_1, \ldots, e_n; e_1 \cdot e_2, e_1 \cdot e_3, \ldots, e_{n-1} \cdot e_n; \ldots; e_1 \cdots e_n\},$$

from which, on account of

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n,$$

we can easily read off the dimension of $C(V, \langle \cdot, \cdot \rangle)$.

If $n = 0$, we obtain $C(V, \langle \cdot, \cdot \rangle) \cong \mathbb{R}$. For $n = 1$, we have $C(V, \langle \cdot, \cdot \rangle) \cong \mathbb{C}$. Namely, if V is generated by $i := e_1$, then $C(V, \langle \cdot, \cdot \rangle)$ is generated by 1 and i with $i^2 = -1$. For $n = 2$, we obtain $C(V, \langle \cdot, \cdot \rangle) \cong \mathbb{H}$; namely, if V is generated

by $i := e_1$ and $j := e_2$, then $C(V, \langle \cdot, \cdot \rangle)$ will be generated by $1, i, j$, and $k := e_1 \cdot e_2$, where $i^2 = j^2 = k^2 = -1$ and $i \cdot j = k = -j \cdot i$. For $n = 3$, one obtains, however, $C(V, \langle \cdot, \cdot \rangle) \cong \mathbb{H} \times \mathbb{H}$; associativity is preserved, but one loses the property of being a division algebra, since now $C(V, \langle \cdot, \cdot \rangle)$ is no longer free of zero divisors.

The Clifford algebra $C(V, \langle \cdot, \cdot \rangle)$ associated with an n-dimensional Euclidean vector space $(V, \langle \cdot, \cdot \rangle)$ makes it possible to generalize the relationship described in Section 3 of this chapter between the special orthogonal group $\mathrm{SO}_3(\mathbb{R})$ and the quaternions $\mathbb{S}^3$ of modulus 1 with the help of the *spinor representation* of the special orthogonal group $\mathrm{SO}_n(\mathbb{R})$ in the group of invertible elements of $C(V, \langle \cdot, \cdot \rangle)$.

References

[1] J. C. Baez: *The octonions.* Bull. Amer. Math. Soc. (N.S.) **39** (2002), 145–205.

[2] J. C. Baez: *Errata for: "The octonions"* [Bull. Amer. Math. Soc. (N.S.) **39** (2002), 145–205]. Bull. Amer. Math. Soc. (N.S.) **42** (2005), 213.

[3] J. H. Conway and D. A. Smith: *On quaternions and octonions: their geometry, arithmetic, and symmetry.* A K Peters, Natick, MA, 2003.

[4] H. Ebbinghaus et al.: *Numbers.* Translated from the 2nd German 1988 edition by H. L. S. Orde. Springer, Berlin Heidelberg New York, 1991.

[5] H. Hopf: *Ein topologischer Beitrag zur reellen Algebra.* Comm. Math. Helvetici **13** (1940/41), 427–440.

[6] A. Hurwitz: *Über die Komposition der quadratischen Formen.* Math. Ann. **88** (1922), 1–25.

[7] J. Milnor: *Some consequences of a theorem of Bott.* Ann. of Math. (2) **68** (1958), 444–449.

[8] B. L. van der Waerden: *A history of algebra. From al-Khwarizmi to Emmy Noether.* Springer, Berlin Heidelberg New York, 1985.

Solutions to Exercises

In this chapter we present solutions to some exercises along with hints for helping to solve others.

Solutions to Exercises in Chapter I

Exercise 1.8. We prove first the validity of the associative law of addition, that is, the equality

$$n + (m + p) = (n + m) + p \tag{1}$$

for all $n, m, p \in \mathbb{N}$. We do this by induction on p. For $p = 0$, the assertion is clear. Suppose, then, that assertion (1) holds for an arbitrary but fixed $p \in \mathbb{N}$ and for all $n, m \in \mathbb{N}$. Then from the definition of addition and the induction hypothesis, we have the equality

$$\begin{aligned}(n + m) + p^* &= ((n + m) + p)^* = (n + (m + p))^* \\ &= n + (m + p)^* = n + (m + p^*),\end{aligned}$$

as desired. This proves the associativity of addition. The first distributive law, that is, the equality

$$(n + m) \cdot p = n \cdot p + m \cdot p \tag{2}$$

for all $n, m, p \in \mathbb{N}$, can be proved using the associativity and commutativity of addition along with induction on p. For $p = 0$, the assertion is clear. Suppose now that (2) holds for an arbitrary but fixed $p \in \mathbb{N}$ and for all $n, m \in \mathbb{N}$. Then we have

$$\begin{aligned}(n + m) \cdot p^* &= (n + m) \cdot p + (n + m) = (n \cdot p + m \cdot p) + (n + m) \\ &= (n \cdot p + n) + (m \cdot p + m) = n \cdot p^* + m \cdot p^*,\end{aligned}$$

as desired. This proves the asserted distributivity. Using the first distributive law, one can then prove the commutativity of multiplication by induction. This then also yields the validity of the second distributive law. Finally, one can prove the associativity of multiplication by induction.

Exercise 1.10. Let $m, n \in \mathbb{N}$. If $m = 0$ or $n = 0$, then one immediately has the equality $m \cdot n = 0$ by Definition 1.5 (2) and the commutativity of mul-

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0

tiplication. To prove the converse, we assume $m \neq 0$ and $n \neq 0$ and prove the inequality $m \cdot n \neq 0$. Since $m \neq 0$ and $n \neq 0$, there exist $a, b \in \mathbb{N}$ with $m = a^* = a + 1$ and $n = b^* = b + 1$. Therefore,

$$m \cdot n = m \cdot b^* = (m \cdot b) + m = (m \cdot b) + (a + 1) = (m \cdot b + a) + 1 = (m \cdot b + a)^*,$$

that is, the natural number $m \cdot n$ is the successor of the natural number $m \cdot b + a$. By the third Peano axiom, we must then have $m \cdot n \neq 0$.

Exercise 1.14. The power law from Lemma 1.13 can by proved by induction.

Exercise 1.17. The proof of properties (i), (ii), and (iii) of Remark 1.16 are left to the reader.

Exercise 1.20. Properties (i) and (ii) of Remark 1.19 can be proved by induction.

Exercise 1.23. We leave it to the reader to come up with suitable examples.

Exercise 1.25. Let $m, n \in \mathbb{N}$ with $m \geq n$. We first prove the existence of a natural number $x \in \mathbb{N}$ with $n + x = m$. If $m = n$, then for $x = 0$, we have the equality $n + x = n + 0 = m$. If $m > n$, then there exists $a \in \mathbb{N}$, $a > 0$, such that $m = n^{*\cdots*}$ (a times). Then for $x = 0^{*\cdots*}$ (a times), we have the equality $n + x = n + 0^{*\cdots*} = n^{*\cdots*} = m$. To prove uniqueness, let $y \in \mathbb{N}$ be another natural number with $n + y = m$. If $x < y$, then by Remark 1.19 (i) and the commutativity of addition, we have the inequality

$$m = n + x = x + n < y + n = n + y = m,$$

that is, we conclude that $m < m$, a contradiction. If $x > y$, then there follows analogously the contradiction $m > m$. Therefore, we must have $x = y$, proving the asserted uniqueness.

Exercise 2.5. (a) By assumption we have $3 \mid (a_1 \cdots a_k + 1)$, that is, there exists $n \in \mathbb{N}$ with $a_1 \cdots a_k + 1 = 3 \cdot n$. If now $3 \mid a_j$ for some $j \in \{1, \ldots, k\}$, then by Lemma 2.4 (ix), we have $3 \mid (a_1 \cdots a_k)$, that is, there exists $m \in \mathbb{N}$ with $a_1 \cdots a_k = 3 \cdot m$. We therefore have the equality

$$1 = 3 \cdot n - a_1 \cdots a_k = 3 \cdot n - 3 \cdot m = 3 \cdot (n - m),$$

a contradiction. Thus none of the numbers $a_1, \ldots, a_k$ is divisible by 3.

(b) We assume that none of the numbers $a_1 + 1, \ldots, a_k + 1$ is divisible by 3. One first shows that then one must have $a_j + 1 = 3 \cdot n_j + r_j$ for certain $n_j \in \mathbb{N}$ and $r_j \in \{1, 2\}$ ($j = 1, \ldots, k$), from which follows $a_j = 3 \cdot n_j + (r_j - 1)$, which for $j = 1, \ldots, k$ implies the equality $r_j = 2$, since by part (a), no a_j is divisible by 3. By multiplying out, one shows that the number

$$a_1 \cdots a_k - 1 = \prod_{j=1}^{k} (3 \cdot n_j + 1) - 1$$

is divisible by 3. But this contradicts the assumption that $a_1 \cdots a_k + 1$ is divisible by 3. Therefore, at least one of the numbers $a_1 + 1, \ldots, a_k + 1$ must be divisible by 3.

Exercise 2.12. To make the proceedings clear, we calculate first, with $a_1 = 2$ and $a_{n+1} = (a_n - 1) \cdot a_n + 1$ for $n \in \mathbb{N}$, $n \geq 1$, the numbers $a_2 = 3$, $a_3 = 7$, $a_4 = 43$, $a_5 = 1807$, $a_6 = (1807 - 1) \cdot 1806 + 1 = 3263443$. With $\mathcal{M}_n = \{p \in \mathbb{P} \mid p \mid a_n\}$, we obtain the first six sets,

$$\mathcal{M}_1 = \{2\}, \quad \mathcal{M}_2 = \{3\}, \quad \mathcal{M}_3 = \{7\}, \quad \mathcal{M}_4 = \{43\},$$
$$\mathcal{M}_5 = \{13, 139\}, \quad \mathcal{M}_6 = \{3263443\}.$$

We claim that we have the equality $a_n = 5 \cdot b_n + r_n$ with $b_n \in \mathbb{N}$ and $r_n \in \{2,3\}$ ($n \in \mathbb{N}$, $n \geq 1$). This can be proved by induction on n. If $n = 1$, the assertion is clear. We now assume that the assertion holds for arbitrary but fixed $n \in \mathbb{N}$, $n \geq 1$. Then we have

$$\begin{aligned} a_{n+1} &= (a_n - 1) \cdot a_n + 1 = (5 \cdot b_n + r_n - 1) \cdot (5 \cdot b_n + r_n) + 1 \\ &= 5 \cdot (5b_n^2 + 2b_n r_n - b_n) + r_n^2 - r_n + 1. \end{aligned}$$

If $r_n = 2$, then $r_n^2 - r_n + 1 = 3$; if $r_n = 3$, then $r_n^2 - r_n + 1 = 5 + 2$. Therefore, in both cases, a_{n+1} is of the form $5 \cdot b_{n+1} + r_{n+1}$ with $b_{n+1} \in \mathbb{N}$ and $r_{n+1} \in \{2,3\}$, as desired. We have thus shown that $5 \nmid a_n$, that is, $5 \notin \mathcal{M}_n$, for all $n \in \mathbb{N}$, $n \geq 1$.

Exercise 2.13. Let us assume, contrary to the assertion, that there are only finitely many prime numbers $p_1, \ldots, p_n$ in $2 + 3 \cdot \mathbb{N}$. We then consider the natural number

$$a := 3 \cdot p_1 \cdots p_n - 1.$$

We have that $a > 1$, and by Lemma 2.9, a has a prime divisor p. Since $3 \nmid a$, it follows that $p \neq 3$. We now show that $p \in 2 + 3 \cdot \mathbb{N}$; that is, we must show that $3 \mid (p + 1)$. If $p = a$, we are done. If $p < a$, then there exists $q \in \mathbb{N}$, $q > 1$, with $a = p \cdot q$. Since $3 \mid (p \cdot q + 1)$, it follows by Exercise 2.5 (b) that $3 \mid (p + 1)$ or $3 \mid (q + 1)$. In the first case, we are done. In the second case, we repeat the process for a prime divisor of q. Finally, after finitely many steps, we obtain a prime divisor p of a with $p \in 2 + 3 \cdot \mathbb{N}$. We now proceed as in Euclid's proof, for on the assumption that there are only finitely many prime numbers in the set $2 + 3 \cdot \mathbb{N}$, we must have $p \in \{p_1, \ldots, p_n\}$. In particular, we have $p \mid (p_1 \cdots p_n)$. However, since we have the divisibility relation $p \mid a$, we must have $p \mid 1$ from the divisibility rules, which is a contradiction.

Exercise 2.15. We prove the contrapositive of the asserted implication.

(i) Suppose that n is not a prime. Then there exist natural numbers $a,b \in \mathbb{N}$ with $n = a \cdot b$ and $1 < a,b < n$. We thus obtain

$$2^n - 1 = 2^{a\cdot b} - 1 = (2^a - 1) \cdot (2^{a\cdot(b-1)} + 2^{a\cdot(b-2)} + \cdots + 2^a + 1).$$

Since $1 < 2^a - 1 < 2^n - 1$, it follows that $2^a - 1$ is a nontrivial divisor of $2^n - 1$, which proves that $2^n - 1$ is not prime.

(ii) Let $n \in \mathbb{N}$, $n > 0$, with n not a power of 2. Then $n > 2$, and there exist natural numbers $a,b \in \mathbb{N}$, b odd, with $n = a \cdot b$ and $1 \leq a < n$, $1 < b \leq n$. Since b is odd, we obtain

$$2^n + 1 = 2^{a\cdot b} + 1 = (2^a + 1) \cdot (2^{a\cdot(b-1)} \mp \cdots - 2^a + 1).$$

Since $1 < 2^a + 1 < 2^n + 1$, it follows that $2^a + 1$ is a nontrivial divisor of $2^n + 1$, which proves that $2^n + 1$ is not prime.

Exercise 2.18. Let $\{a_1,\ldots,a_n\}$ and $\{b_1,\ldots,b_m\}$ denote the sets of all divisors of a and all divisors of b. Since a and b are relatively prime, the set of all divisors of $a \cdot b$ is equal to $\{a_j \cdot b_k \mid j = 1,\ldots,n; k = 1,\ldots,m\}$. It follows that

$$S(a) \cdot S(b) = (a_1 + \cdots + a_n) \cdot (b_1 + \cdots + b_m) = \sum_{j=1}^{n} \sum_{k=1}^{m} a_j \cdot b_k = S(a \cdot b).$$

This completes the proof of the assertion.

Exercise 2.19. The assertion can be proved by induction on m.

Exercise 2.20. (a) We have the equalities

$$S(220) - 220 = 1 + 2 + 4 + 5 + 10 + 11 + 20 + 22 + 44 + 55 + 110 = 284,$$
$$S(284) - 284 = 1 + 2 + 3 + 71 + 142 = 220.$$

Therefore, we have $S(220) = 220 + 284 = S(284)$, which proves that the numbers 220 and 284 are amicable.

(b) We must show that $S(a) = a + b = S(b)$. Since x,y,z are distinct odd primes, it follows that

$$S(a) = S(2^n \cdot x \cdot y) = S(2^n) \cdot S(x) \cdot S(y) = (2^{n+1} - 1)(x + 1)(y + 1),$$
$$S(b) = S(2^n) \cdot S(z) = (2^{n+1} - 1)(z + 1),$$

where we have used the well-known equality $S(2^n) = 2^{n+1} - 1$. A direct calculation shows that $x \cdot y = 9 \cdot 2^{2n-1} - 9 \cdot 2^{n-1} + 1$, and therefore, $x \cdot y + x + y = z$. This implies $S(a) = (2^{n+1} - 1)(z + 1) = S(b)$. Finally, we calculate

$$a+b = 2^n\cdot(x\cdot y+z) = 2^n\cdot(9\cdot 2^{2n}-9\cdot 2^{n-1}) = 2^{2n-1}\cdot 9\cdot(2^{n+1}-1)$$
$$= (z+1)\cdot(2^{n+1}-1) = S(a) = S(b),$$

which proves that the numbers a and b are amicable.

Exercise 3.2. We obtain the prime factorizations $720 = 2^4\cdot 3^2\cdot 5$, $9\,797 = 97\cdot 101$ and $360^{360} = (2^3\cdot 3^2\cdot 5)^{360} = 2^{1\,080}\cdot 3^{720}\cdot 5^{360}$. Finally, on using the third binomial formula four times, we obtain

$$2^{32}-1 = (2^2-1)\cdot(2^2+1)\cdot(2^4+1)\cdot(2^8+1)\cdot(2^{16}+1)$$
$$= 3\cdot 5\cdot 17\cdot 257\cdot 65\,537.$$

Exercise 3.7. Let $a = 2^{32}-1$ and $b = 255$ with prime factorizations (see Exercise 3.2)

$$a = \prod_{p\in\mathbb{P}} p^{a_p} = 3\cdot 5\cdot 17\cdot 257\cdot 65\,537, \quad b = \prod_{p\in\mathbb{P}} p^{b_p} = 3\cdot 5\cdot 17;$$

here $a_3 = 1$, $a_5 = 1$, $a_{17} = 1$, $a_{257} = 1$, $a_{65\,537} = 1$, $a_p = 0$ for all $p \in \mathbb{P}\setminus\{3,5,17,257,65\,537\}$ and $b_3 = 1, b_5 = 1, b_{17} = 1, b_p = 0$ for all $p\in\mathbb{P}\setminus\{3,5,17\}$. Therefore, $b_p \le a_p$ for all $p\in\mathbb{P}$, which by the criterion of Lemma 3.5, proves that $b \mid a$.

Exercise 4.6. With the help of Theorem 4.3, we obtain $(3\,600, 3\,240) = 360$, $(360^{360}, 540^{180}) = ((2^3\cdot 3^2\cdot 5)^{360}, (2^2\cdot 3^3\cdot 5)^{180}) = 2^{360}\cdot 3^{540}\cdot 5^{180}$, $(2^{32}-1, 3^8-2^8) = 5$, where for the last equality, we used the prime factorization from Exercise 3.2 and the prime factorization $3^8-2^8 = (3^2-2^2)\cdot(3^2+2^2)\cdot(3^4+2^4) = 5\cdot 13\cdot 97$.

Exercise 4.13. We have $(2\,880, 3\,000, 3\,240) = (120, 3\,240) = 120$ and $[36, 42, 49] = [252, 49] = 1\,764$.

Exercise 4.15. For example, the numbers $a_1 = 6$, $a_2 = 10$, $a_3 = 15$ are relatively prime, since $(a_1, a_2, a_3) = (6, 10, 15) = (2, 15) = 1$. The numbers a_1, a_2, a_3, however, are not pairwise relatively prime, since we have $(a_1, a_2) = 2$.

Exercise 4.17. Let $a_1, \ldots, a_n \in \mathbb{N}$. We leave to the reader the proof of the following equivalence:

$$(a_1,\ldots,a_n)\cdot[a_1,\ldots,a_n] = a_1\cdots a_n \iff a_1,\ldots,a_n \text{ pairwise relatively prime.}$$

This proves the desired criterion.

Exercise 5.2. We obtain $773 = 2\cdot 337 + 99$. Further, we calculate $2^5\cdot 3^4\cdot 5^2 = (2^2\cdot 3^2)\cdot(2^3\cdot 3^2\cdot 5^2) = (5\cdot 7+1)\cdot(2^3\cdot 3^2\cdot 5^2) = 7\cdot(2^3\cdot 3^2\cdot 5^3) + (2^3\cdot 3^2\cdot 5^2)$. Since $2^{16}+1 = 4^8+1$, it follows that $2^{32}-1 = (2^{16}-1)(4^8+1)+0$.

Exercise 5.4. This process can be carried out for arbitrary natural numbers $g > 1$. One obtains the unique representation

$$n = q_\ell \cdot g^\ell + q_{\ell-1} \cdot g^{\ell-1} + \cdots + q_1 \cdot g^1 + q_0 \cdot g^0$$

with natural numbers $0 \le q_j \le g-1$ $(j = 0, \ldots, \ell)$ and $q_\ell \neq 0$, called the *g-adic representation* of the natural number n.

Solutions to Exercises in Chapter II

Exercise 1.3. For $a, b, c \in \mathcal{R}_n$, one has the equalities

$$\begin{aligned}(a \oplus b) \oplus c &= \mathcal{R}_n(a+b) \oplus c = \mathcal{R}_n(\mathcal{R}_n(a+b)+c),\\ a \oplus (b \oplus c) &= a \oplus \mathcal{R}_n(b+c) = \mathcal{R}_n(a+\mathcal{R}_n(b+c)).\end{aligned}$$

Division with remainder of $a+b$ and $b+c$ by n yields the uniquely determined numbers $q_1, q_2 \in \mathbb{N}$ such that

$$a+b = q_1 \cdot n + \mathcal{R}_n(a+b) \quad \text{and} \quad b+c = q_2 \cdot n + \mathcal{R}_n(b+c),$$

whence follows

$$\begin{aligned}\mathcal{R}_n(\mathcal{R}_n(a+b)+c) &= \mathcal{R}_n(a+b+c-q_1 \cdot n) = \mathcal{R}_n(a+b+c)\\ &= \mathcal{R}_n(a+b+c-q_2 \cdot n) = \mathcal{R}_n(a+\mathcal{R}_n(b+c)).\end{aligned}$$

We have thereby shown that the operation $\oplus$ is associative. Analogously, one can prove that the operation $\odot$ is associative.

Exercise 1.4. (a) The set $2 \cdot \mathbb{N} = \{2 \cdot n \mid n \in \mathbb{N}\}$ of even natural numbers is a nonempty subset of $\mathbb{N}$. If $2 \cdot m, 2 \cdot n \in 2 \cdot \mathbb{N}$, then

$$2 \cdot m + 2 \cdot n = 2 \cdot (m+n) \in 2 \cdot \mathbb{N}, \quad (2 \cdot m) \cdot (2 \cdot n) = 2 \cdot (m \cdot 2 \cdot n) \in 2 \cdot \mathbb{N}.$$

Thus both $+$ and $\cdot$ are operations on $2 \cdot \mathbb{N}$. Since the operations $+$ on $\mathbb{N}$ and $\cdot$ on $\mathbb{N}$ are associative, it follows that in particular, the operations $+$ on $2 \cdot \mathbb{N}$ and $\cdot$ on $2 \cdot \mathbb{N}$ are associative. Therefore, both $(2 \cdot \mathbb{N}, +)$ and $(2 \cdot \mathbb{N}, \cdot)$ are semigroups.

The set $2 \cdot \mathbb{N} + 1 = \{2 \cdot n + 1 \mid n \in \mathbb{N}\}$ of odd natural numbers is a nonempty subset of $\mathbb{N}$. If $2 \cdot m + 1$ and $2 \cdot n + 1$ are in $2 \cdot \mathbb{N} + 1$, then

$$\begin{aligned}(2 \cdot m + 1) + (2 \cdot n + 1) &= 2 \cdot (m + n + 1) \in 2 \cdot \mathbb{N},\\ (2 \cdot m + 1) \cdot (2 \cdot n + 1) &= 2 \cdot (m \cdot 2 \cdot n + m + n) + 1 \in 2 \cdot \mathbb{N} + 1.\end{aligned}$$

Therefore, while $\cdot$ is an operation on $2 \cdot \mathbb{N} + 1$, we see that $+$ is not an operation on $2 \cdot \mathbb{N} + 1$. Therefore, $(2 \cdot \mathbb{N} + 1, +)$ is not a semigroup. The operation $\cdot$ on $\mathbb{N}$ is associative, and so in particular, the operation $\cdot$ on $2 \cdot \mathbb{N} + 1$ is associative. Therefore, $(2 \cdot \mathbb{N} + 1, \cdot)$ is a semigroup.

(b) Let $k \in \mathbb{N}$, $k > 1$. The set $k \cdot \mathbb{N} = \{k \cdot n \mid n \in \mathbb{N}\}$ is a nonempty subset of $\mathbb{N}$. One shows, as in (a), that both $(k \cdot \mathbb{N}, +)$ and $(k \cdot \mathbb{N}, \cdot)$ are semigroups.

Exercise 1.5. Because of the inequality

$$(2 \circ 3) \circ 2 = (2^3) \circ 2 = (2^3)^2 = 2^{3 \cdot 2} = 2^6 \neq 2^9 = 2 \circ (3^2) = 2 \circ (3 \circ 2),$$

the operation $\circ$ on $\mathbb{N}$ is not associative, and therefore $(\mathbb{N}, \circ)$ is not a semigroup.

Exercise 1.8. If $A_1 = \{a_1\}$ is a one-element set, then

$$\mathrm{map}(A_1) = \{\mathrm{id}\},$$

where the mapping $\mathrm{id} : A_1 \longrightarrow A_1$ is given by the assignment $a_1 \mapsto a_1$. The semigroup $(\mathrm{map}(A_1), \circ)$ is abelian. If $A_2 = \{a_1, a_2, \dots\}$ is an arbitrary set that contains at least two elements $a_1 \neq a_2$, then

$$\mathrm{map}(A_2) = \{\mathrm{id}, f, g, \dots\},$$

where the mapping $\mathrm{id} : A_2 \longrightarrow A_2$ is given by $a_j \mapsto a_j$ $(a_j \in A_2)$, the mapping $f : A_2 \longrightarrow A_2$ by $a_1 \mapsto a_2$, $a_j \mapsto a_j$ $(a_j \in A_2 \setminus \{a_1\})$, and the mapping $g : A_2 \longrightarrow A_2$ by $a_2 \mapsto a_1$, $a_j \mapsto a_j$ $(a_j \in A_2 \setminus \{a_2\})$. But then we have

$$(f \circ g)(a_1) = f(g(a_1)) = f(a_1) = a_2 \neq a_1 = g(a_2) = g(f(a_1)) = (g \circ f)(a_1),$$

whence $(\mathrm{map}(A_2), \circ)$ is a nonabelian semigroup.

Exercise 1.12. Let e_ℓ be a left identity element and e_r a right identity element of H. Then

$$e_\ell = e_\ell \circ e_r = e_r,$$

where the first equality follows from the fact that e_r is a right identity element of H, and the second from the fact that e_ℓ is a left identity element of H.

Exercise 1.14. (a) By Exercise 1.4, $(2 \cdot \mathbb{N}, +)$ and $(2 \cdot \mathbb{N}, \cdot)$ are semigroups. It remains to show that there exists an additive identity element in $2 \cdot \mathbb{N}$. By the definition of addition, 0 is this element. Since $1 \notin 2 \cdot \mathbb{N}$, there is no multiplicative identity element, so that $(2 \cdot \mathbb{N}, \cdot)$ is only a semigroup.

(b) We leave it to the reader to find other examples of semigroups that are not monoids.

Exercise 2.3. (a) Suppose that g' and g'' are two inverse elements to an element $g \in G$. Then

$$g' = g' \circ e = g' \circ (g \circ g'') = (g' \circ g) \circ g'' = e \circ g'' = g'',$$

where the second equality follows from the fact that g'' is in particular a right inverse to g, and the fourth equality follows from the fact that g' is in particular a left inverse to g.
(b) Let g'_ℓ be a left inverse and g'_r a right inverse to an element $g \in G$. Then it follows that

$$g'_\ell = g'_\ell \circ e = g'_\ell \circ (g \circ g'_r) = (g'_\ell \circ g) \circ g'_r = e \circ g'_r = g'_r,$$

analogously to part (a).

Exercise 2.6. (a) Let $g^{-1} \in G$ be the inverse element to $g \in G$. Then

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

Thus g is the inverse element to g^{-1}, that is, $(g^{-1})^{-1} = g$.
(b) Let $g^{-1} \in G$ be the inverse element to $g \in G$ and $h^{-1} \in G$ the inverse element to $h \in G$. Then

$$(h^{-1} \circ g^{-1}) \circ (g \circ h) = h^{-1} \circ (g^{-1} \circ g) \circ h = h^{-1} \circ e \circ h = h^{-1} \circ h = e,$$
$$(g \circ h) \circ (h^{-1} \circ g^{-1}) = g \circ (h \circ h^{-1}) \circ g^{-1} = g \circ e \circ g^{-1} = g \circ g^{-1} = e.$$

Thus $h^{-1} \circ g^{-1}$ is the inverse element to $g \circ h$, that is, $(g \circ h)^{-1} = h^{-1} \circ g^{-1}$.
The calculational rules (c) and (d) follow directly from the definition.

Exercise 2.9. We compare only the groups that have the same numbers of elements. The Cayley tables of the groups $(\mathcal{R}_4, \oplus)$, $(\mathcal{R}_5 \setminus \{0\}, \odot)$, and $(D_4, \circ)$ have, reading from left to right, the following form:

$\oplus$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\odot$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

$\circ$	d_0	d_1	s_0	s_1
d_0	d_0	d_1	s_0	s_1
d_1	d_1	d_0	s_1	s_0
s_0	s_0	s_1	d_0	d_1
s_1	s_1	s_0	d_1	d_0

One may conclude from the Cayley tables that all three groups under consideration are abelian. We now determine the smallest nonzero natural number n such that $g^n = e$ for $g \neq e$. In the group $(\mathcal{R}_4, \oplus)$, we have $e = 0$ and

$$1^2 = 2, 1^3 = 3, 1^4 = 0; \quad 2^2 = 0; \quad 3^2 = 2, 3^3 = 1, 3^4 = 0.$$

In the group $(\mathcal{R}_5 \setminus \{0\}, \odot)$, we have $e = 1$ and

$$2^2 = 4, 2^3 = 3, 2^4 = 1; \quad 3^2 = 4, 3^3 = 2, 3^4 = 1; \quad 4^2 = 1.$$

Thus in each group there are two elements with $n = 4$ and one element with $n = 2$. In the group $(D_4, \circ)$, however, $d_1^2 = s_0^2 = s_1^2 = e$ with $e = d_0$, that is, there is no element with $n = 4$.
The Cayley tables for $(\mathcal{R}_6, \oplus)$ and $(D_6, \circ)$, reading from left to right, have the following form:

$\oplus$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\circ$	d_0	d_1	d_2	s_0	s_1	s_2
d_0	d_0	d_1	d_2	s_0	s_1	s_2
d_1	d_1	d_2	d_0	s_2	s_0	s_1
d_2	d_2	d_0	d_1	s_1	s_2	s_0
s_0	s_0	s_1	s_2	d_0	d_1	d_2
s_1	s_1	s_2	s_0	d_2	d_0	d_1
s_2	s_2	s_0	s_1	d_1	d_2	d_0

From these tables, one can see that the group $(\mathcal{R}_6, \oplus)$ is abelian. The group $(D_6, \circ)$ is nonabelian, since $s_0 \circ s_1 = d_2 \neq d_1 = s_1 \circ s_0$. We again determine the smallest nonzero natural number n such that $g^n = e$ for $g \neq e$. In $(\mathcal{R}_6, \oplus)$, we have $e = 0$ and

$$1^2 = 2, 1^3 = 3, 1^4 = 4, 1^5 = 5, 1^6 = 0; \quad 2^2 = 4, 2^3 = 0; \quad 3^2 = 0;$$
$$4^2 = 2, 4^3 = 0; \quad 5^2 = 4, 5^3 = 3, 5^4 = 2, 5^5 = 1, 5^6 = 0.$$

There exist, therefore, in $(\mathcal{R}_6, \oplus)$ two elements with $n = 6$, two elements with $n = 3$, and one element with $n = 2$. In $(D_6, \circ)$, we have $e = d_0$ and

$$d_1^2 = d_2,\ d_1^3 = d_0; \quad d_2^2 = d_1,\ d_2^3 = d_0; \quad s_0^2 = d_0,\ s_1^2 = d_0,\ s_2^2 = d_0.$$

There exist, therefore, in $(D_6, \circ)$ no element with $n = 6$, two elements with $n = 3$, and three elements with $n = 2$.

Exercise 2.10. (a) One shows using the Cayley table

$\odot$	1	2
1	1	2
2	2	1

for $(\mathcal{R}_3 \setminus \{0\}, \odot)$ and the Cayley table from Exercise 2.9 for $(\mathcal{R}_5 \setminus \{0\}, \odot)$ that $(\mathcal{R}_3 \setminus \{0\}, \odot)$ and $(\mathcal{R}_5 \setminus \{0\}, \odot)$ are groups.

(b) We leave to the reader the task of verifying the assertions of Example 2.8 (iii) regarding the dihedral group $(D_{2n}, \circ)$.

(c) Let $n \geq 3$. We consider the elements

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 2 & 1 & 3 & \cdots & n \end{pmatrix}, \quad \pi_2 = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 3 & 1 & 2 & \cdots & n \end{pmatrix}$$

of S_n, where for $n > 3$, each of the elements $4, \ldots, n$ is mapped to itself. Then

$$\pi_1 \circ \pi_2 = \begin{pmatrix} 1\,2\,3\cdots n \\ 3\,2\,1\cdots n \end{pmatrix} \neq \begin{pmatrix} 1\,2\,3\cdots n \\ 1\,3\,2\cdots n \end{pmatrix} = \pi_2 \circ \pi_1,$$

where for $n > 3$, each of the elements $4, \ldots, n$ is mapped to itself. This proves that $(S_n, \circ)$ for $n \geq 3$ is nonabelian.

Exercise 2.13. This is proved by induction on n.

Exercise 2.19. We have $S_3 = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\}$ with

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

We calculate $\operatorname{ord}(\pi_1) = 1$, $\operatorname{ord}(\pi_2) = \operatorname{ord}(\pi_3) = 3$, and $\operatorname{ord}(\pi_4) = \operatorname{ord}(\pi_5) = \operatorname{ord}(\pi_6) = 2$.

Exercise 2.23. Since $d_1^k = d_k$ $(k = 0, \ldots, n-1)$ and $d_1^n = d_0$, we have $\langle d_1 \rangle = \{d_0, \ldots, d_{n-1}\}$. Using the subgroup criterion, one can show that the nonempty subset $\langle g \rangle = \{\ldots, (g^{-1})^2, g^{-1}, g^0 = e, g^1 = g, g^2, \ldots\} \subseteq G$ is a subgroup of G for every group G. In particular, $\langle d_1 \rangle = \{d_0, \ldots, d_{n-1}\}$ is a subgroup of D_{2n}.

Exercise 2.26. We have $S_3 = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\}$ with π_j $(j = 1, \ldots, n)$ as in Exercise 2.19. We have the cyclic subgroups

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

and the subgroup S_3 itself, which is not cyclic. One can see that S_3 has no other subgroups.

Exercise 3.3. We have $S_3 = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\}$ with π_j $(j = 1, \ldots, n)$ as in Exercise 2.19 and $D_6 = \{d_0, d_1, d_2, d_0 \circ s_0, d_1 \circ s_0, d_2 \circ s_0\}$, where s_0 is reflection in the median of the side joining vertices 1 and 2. By the definition of the group homomorphism $f : D_6 \longrightarrow S_3$, we have

$$f(d_0) = \pi_1, f(d_1) = \pi_3, f(d_2) = \pi_2,$$
$$f(d_0 \circ s_0) = \pi_6, f(d_1 \circ s_0) = \pi_5, f(d_2 \circ s_0) = \pi_4,$$

which proves that f is bijective and therefore in fact a group isomorphism.

Exercise 3.5. Let $d_{j_1} \circ s_0^{k_1}$ and $d_{j_2} \circ s_0^{k_2}$ with $j_1, j_2 \in \{0, \ldots, n-1\}$ and $k_1, k_2 \in \{0, 1\}$ be two elements of D_{2n}. Since $d_j \circ s_0 = s_0 \circ d_j^{-1}$, we have

$$(d_{j_1} \circ s_0^{k_1}) \circ (d_{j_2} \circ s_0^{k_2}) = \begin{cases} d_{j_1} \circ d_{j_2}, & \text{if } k_1 = 0, k_2 = 0; \\ d_{j_1} \circ d_{j_2} \circ s_0, & \text{if } k_1 = 0, k_2 = 1; \\ d_{j_1} \circ d_{j_2}^{-1}, & \text{if } k_1 = 1, k_2 = 1; \\ d_{j_1} \circ d_{j_2}^{-1} \circ s_0, & \text{if } k_1 = 1, k_2 = 0. \end{cases}$$

It follows that

$$\operatorname{sgn}\big((d_{j_1} \circ s_0^{k_1}) \circ (d_{j_2} \circ s_0^{k_2})\big) = k_1 \oplus k_2 = \operatorname{sgn}(d_{j_1} \circ s_0^{k_1}) \oplus \operatorname{sgn}(d_{j_2} \circ s_0^{k_2}).$$

Therefore, sgn is a group homomorphism, and we have $\operatorname{im}(\operatorname{sgn}) = \mathcal{R}_2$ and $\ker(\operatorname{sgn}) = \{d_j \mid j = 0, \ldots, n-1\}$.

Exercise 3.7. By Lemma 3.6, we have f injective $\iff \ker(f) = \{e_G\}$. It therefore suffices to prove, under the assumption $|G| < \infty$, the equivalence f injective $\iff f$ surjective. We have

$$\begin{aligned} f \text{ injective } &\iff g \neq h \text{ for } g, h \in G \text{ implies } f(g) \neq f(h) \\ &\underset{|G|<\infty}{\iff} |f(G)| = |G| \\ &\iff \text{for every } g \in G \text{ there exists } h \in G \text{ with } f(h) = g \\ &\iff f \text{ surjective.} \end{aligned}$$

This proves the assertion.

Exercise 3.8. If $g \in G$ and $\operatorname{ord}(g) = n$, then $e = f(e) = f(g^n) = f(g)^n$, which implies $\operatorname{ord}\big(f(g)\big) \leq n = \operatorname{ord}(g)$.

Exercise 3.9. Suppose that $f : D_{24} \longrightarrow S_4$ is a group isomorphism. Then for each $g \in D_{24}$, we must have the equality

$$\operatorname{ord}(g) = \operatorname{ord}(f(g)).$$

Since $\operatorname{ord}(d_2) = 12$, we must have that $f(d_2) \in S_4$ is an element of order 12. We first determine the orders of the elements of S_4. We obtain the nine elements of order 2,

$$\begin{pmatrix} 1&2&3&4 \\ 2&1&3&4 \end{pmatrix}, \quad \begin{pmatrix} 1&2&3&4 \\ 3&2&1&4 \end{pmatrix}, \quad \begin{pmatrix} 1&2&3&4 \\ 4&2&3&1 \end{pmatrix},$$
$$\begin{pmatrix} 1&2&3&4 \\ 1&3&2&4 \end{pmatrix}, \quad \begin{pmatrix} 1&2&3&4 \\ 1&4&3&2 \end{pmatrix}, \quad \begin{pmatrix} 1&2&3&4 \\ 1&2&4&3 \end{pmatrix},$$
$$\begin{pmatrix} 1&2&3&4 \\ 2&1&4&3 \end{pmatrix}, \quad \begin{pmatrix} 1&2&3&4 \\ 3&4&1&2 \end{pmatrix}, \quad \begin{pmatrix} 1&2&3&4 \\ 4&3&2&1 \end{pmatrix},$$

the eight elements of order 3,

$$\begin{pmatrix}1&2&3&4\\2&3&1&4\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\3&1&2&4\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\2&4&3&1\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\4&1&3&2\end{pmatrix},$$
$$\begin{pmatrix}1&2&3&4\\3&2&4&1\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\4&2&1&3\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\1&3&4&2\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\1&4&2&3\end{pmatrix},$$

and the six elements of order 4,

$$\begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\2&4&1&3\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\3&4&2&1\end{pmatrix},$$
$$\begin{pmatrix}1&2&3&4\\3&1&4&2\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\4&3&1&2\end{pmatrix},\quad \begin{pmatrix}1&2&3&4\\4&1&2&3\end{pmatrix}.$$

The group S_4 has, therefore, only elements of orders 1, 2, 3, and 4. There can therefore be no group isomorphism between D_{24} and S_4.

Exercise 3.11. (a) If $f : (\mathcal{R}_4, \oplus) \longrightarrow (\mathcal{R}_4, \oplus)$ is a group homomorphism, then $f(0) = 0$. Since $\mathcal{R}_4 = \langle 1 \rangle$, it follows that

$$f(2) = f(1 \oplus 1) = f(1) \oplus f(1), f(3) = f(1 \oplus 1 \oplus 1) = f(1) \oplus f(1) \oplus f(1),$$

that is, f is uniquely determined by specifying the image of $f(1)$. Therefore, there are precisely four distinct group homomorphisms, f_1, f_2, f_3, f_4, given by the assignments

$$\begin{aligned}
f_1(0) = 0, f_1(1) = 0, f_1(2) = 0, f_1(3) = 0,\\
f_2(0) = 0, f_2(1) = 1, f_2(2) = 2, f_2(3) = 3,\\
f_3(0) = 0, f_3(1) = 2, f_3(2) = 0, f_3(3) = 2,\\
f_4(0) = 0, f_4(1) = 3, f_4(2) = 2, f_4(3) = 1,
\end{aligned}$$

whence follows $\ker(f_1) = \mathcal{R}_4$, $\ker(f_2) = \{0\}$, $\ker(f_3) = \{0,2\}$, $\ker(f_4) = \{0\}$, $\mathrm{im}(f_1) = \{0\}$, $\mathrm{im}(f_2) = \mathcal{R}_4$, $\mathrm{im}(f_3) = \{0,2\}$, $\mathrm{im}(f_4) = \mathcal{R}_4$. This shows in particular that f_2 and f_4 are bijective.

(b) Since $\mathcal{R}_p = \langle 1 \rangle$, every group homomorphism $g : (\mathcal{R}_p, \oplus) \longrightarrow (\mathcal{R}_n, \oplus)$ is uniquely determined by specifying the image $g(1)$. One now shows that only $g(1) = 0$ is possible, since n and p are relatively prime. There is, therefore, only one group homomorphism g, which is given by the assignment $g(m) = 0$ ($m \in \mathcal{R}_p$). We have $\ker(g) = \mathcal{R}_p$ and $\mathrm{im}(g) = \{0\}$.

Exercise 4.3. (a) The verification of the statement of Example 4.2 is left to the reader.

(b) The order relation $\leq$ is not an equivalence relation on $\mathbb{N}$, since $\leq$ is not symmetric.

(c) The relation $\sim$ is not an equivalence relation on $\mathbb{N}$, since $\sim$ is not transitive.

Exercise 4.6. Let M be a set with an element $m \in M$. The equivalence class of m with respect to equality "$=$" is $M_m = \{m' \in M \mid m' = m\}$, that is, the set of all elements of M that are equal to m. We leave it to the reader to find additional equivalence relations and determine the associated equivalence classes.

Exercise 4.11. We leave the solution of this exercise to the reader.

Exercise 4.12. Let $U = \langle \pi_4 \rangle = \{\pi_1, \pi_4\}$. The left coset of an element $\pi \in S_3$ with respect to U is given by $\pi \circ U = \{\pi \circ \pi_1, \pi \circ \pi_4\}$. Therefore, the following are all left cosets of S_3 with respect to U:

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

Exercise 4.15. (a) If $g \in G$, then $\operatorname{ord}(g) = |U|$ for $U = \langle g \rangle \leq G$, which implies by Lagrange's theorem that $\operatorname{ord}(g) \mid |G|$.
(b) Suppose $|G| = p$ for a prime number p. If $g \in G$, $g \neq e$, then $\operatorname{ord}(g) > 1$, and therefore, by part (a), we must have the equality $\operatorname{ord}(g) = p$, which implies $G = \langle g \rangle$.
(c) Let $|G| = 4$, and write $G = \{e, a, b, c\}$. If G has an element $g \in G$ with $\operatorname{ord}(g) = 4$, then $G = \langle g \rangle$, that is, G is cyclic, and therefore isomorphic to the group $(\mathcal{R}_4, \oplus)$. If G has no element of order 4, then every element $g \in G$, $g \neq e$, has order 2, that is, $a^2 = b^2 = c^2 = e$. Since G is a group, it follows that $a \circ b = c = b \circ a$, $a \circ c = b = c \circ a$, and $b \circ c = a = c \circ b$. Therefore, G is isomorphic to the group $(D_4, \circ)$. Therefore, up to group isomorphism, there are precisely two groups of order 4, given by the following Cayley tables:

$\circ$	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

$\circ$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Both groups are abelian.
One shows further that up to group isomorphism, the only groups of order 6 are $(\mathcal{R}_6, \oplus)$ and $(D_6, \circ)$. Thus every abelian group of order 6 is isomorphic to $(\mathcal{R}_6, \oplus)$, and every nonabelian group of order 6 is isomorphic to $(D_6, \circ)$.

Exercise 4.19. Exercises 4.11 and 4.12 can be solved analogously for right cosets.

Exercise 4.21. Let $U = \langle \pi_4 \rangle = \{\pi_1, \pi_4\}$. Then $U \circ \pi_2 = \{\pi_2, \pi_5\} \neq \{\pi_2, \pi_6\} = \pi_2 \circ U$, which implies that U is not a normal subgroup of S_3. Analogously, one can show that $\langle \pi_5 \rangle$ and $\langle \pi_6 \rangle$ are not normal subgroups of S_3.

Exercise 4.24. (a) For every element $h \in H$, one has $h \circ H = H = H \circ h$. Now let $g \in G \setminus H$. Then $g \circ H \neq H$ and $H \circ g \neq H$. Because of $[G : H] = 2$, we obtain the disjoint decomposition $H \dot\cup (g \circ H) = G = H \dot\cup (H \circ g)$ of G. It then follows that $g \circ H = H \circ g$ must hold. Altogether, one has the equality $g \circ H = H \circ g$ for all $g \in G$. This proves that H is a normal subgroup of G.
(b) The mapping $f : G \longrightarrow \mathcal{R}_2$, given by

$$f(g) = \begin{cases} 0, & \text{if } g \in H; \\ 1, & \text{if } g \notin H, \end{cases}$$

is a surjective group homomorphism.

Exercise 4.26. Since $\ker(f)$ is a subgroup, indeed a normal subgroup, of S_3, it must be the case that $\ker(f)$ is one of the groups $\{\pi_1\}$, A_3, S_3. If $\ker(f) = \{\pi_1\}$, then f is injective, which is impossible because of $6 = |S_3| > |\mathcal{R}_3| = 3$. If $\ker(f) = \{A_3\}$, then $\operatorname{ord}(\pi_4) < \operatorname{ord}(f(\pi_4))$, which is impossible because of Exercise 3.8. The only possibility is then $\ker(f) = S_3$, which implies that $f(\pi) = 0$ for all $\pi \in S_3$.

Exercise 5.10. Let $f : G \longrightarrow \mathcal{R}_2$ be the surjective group homomorphism of Exercise 4.24. Then $\ker(f) = H$, and by Corollary 5.8, we have an isomorphism $G/H \cong \mathcal{R}_2$. This isomorphism can also be read off from the Cayley table

$\bullet$	H	$g \circ H$
H	H	$g \circ H$
$g \circ H$	$g \circ H$	H

for the group $G/H = \{H, g \circ H\}$, where $g \in G \setminus H$ is arbitrary, and $H = e_G \circ H$. We note here that $(g \circ H) \bullet (g \circ H) = H$ must hold, since otherwise, we would have from $(g \circ H) \bullet (g \circ H) = (g \circ g) \circ H = g \circ H$ the equalities $g \circ g \circ h_1 = g \circ h_2 \Longleftrightarrow g \circ h_1 = h_2 \Longleftrightarrow g = h_2 \circ h_1^{-1}$ for certain $h_1, h_2 \in H$, and thus the contradiction $g \in H$.

Exercise 5.11. It is clear that from $H \trianglelefteq G$, $K \trianglelefteq G$, and $K \subseteq H$, one has $K \trianglelefteq H$. That H/K is a normal subgroup of G/K comes immediately from the proof of isomorphism. To this end, we define the mapping $f : G/K \longrightarrow G/H$ by the assignment

$$g \circ K \mapsto g \circ H.$$

Since $K \subseteq H$, it follows that f is well defined. On account of $K \trianglelefteq G$ and $H \trianglelefteq G$, it is clear that f is a homomorphism. Furthermore, we have

$$\begin{aligned} \ker(f) &= \{g \circ K \mid f(g \circ K) = H\} = \{g \circ K \mid g \circ H = H\} \\ &= \{g \circ K \mid g \in H\} = H/K. \end{aligned}$$

This proves in particular that $H/K \trianglelefteq G/K$. Since f is surjective, it follows from the homomorphism theorem that

$$(G/K)/(H/K) = (G/K)/\ker(f) \cong G/K,$$

as claimed.

Exercise 6.4. (a) Let $A = \{a_1, a_2, \dots\}$ be a set with $a_1 \neq a_2$. Then $\mathrm{map}(A) = \{\mathrm{id}, f, g, \dots\}$, whereby the mapping $\mathrm{id} : A \longrightarrow A$ is given by the assignment $a_j \mapsto a_j$ $(a_j \in A)$, the mapping $f : A \longrightarrow A$ by $a_1 \mapsto a_2$, $a_j \mapsto a_j$ $(a_j \in A \setminus \{a_1\})$, and the mapping $g : A \longrightarrow A$ by $a_2 \mapsto a_1$, $a_j \mapsto a_j$ $(a_j \in A \setminus \{a_2\})$. It then follows that

$$(f \circ g)(a_1) = f(a_1) = a_2 = (f \circ \mathrm{id})(a_1),\ (f \circ g)(a_2) = f(a_1) = a_2 = (f \circ \mathrm{id})(a_2),$$

which proves $f \circ g = f \circ \mathrm{id}$; however, we have $g \neq \mathrm{id}$, which proves that in the semigroup $(\mathrm{map}(A), \circ)$, the first cancellation law is invalid.

One may show analogously that the second cancellation law is also invalid.

(b) We leave it to the reader to find further examples of semigroups that are not regular.

Exercise 6.6. (a) The solution to this exercise is left to the reader.

(b) On $(2 \cdot \mathbb{N} + 1) \times (2 \cdot \mathbb{N} + 1) = \{(a, b) \mid a, b \in 2 \cdot \mathbb{N} + 1\}$ we define the relation

$$(a, b) \sim (c, d) \Longleftrightarrow a \cdot d = b \cdot c \quad (a, b, c, d \in 2 \cdot \mathbb{N} + 1).$$

If we write $\frac{a}{b}$ for the equivalence class $[a, b]$ of $(a, b) \in (2 \cdot \mathbb{N} + 1) \times (2 \cdot \mathbb{N} + 1)$, then the group $G := ((2 \cdot \mathbb{N} + 1) \times (2 \cdot \mathbb{N} + 1))/\sim$ can be identified with the set of all fractions of the form $\frac{a}{b}$, where $a, b \in 2 \cdot \mathbb{N} + 1$ and a, b are relatively prime. We leave the detailed construction from Theorem 6.5 to the reader.

Exercise 7.6. The generalization to $\mathbb{Z}$ of the addition and multiplication rules in Remark 1.19 of Chapter I is left to the reader.

Exercise 7.9. The verification of the assertions of this example are left to the reader.

Solutions to Exercises in Chapter III

Exercise 1.2. The proof of the calculational laws from Lemma 1.1 are left to the reader.

Exercise 1.5. The proof of Theorem 1.4 is left to the reader.

Exercise 1.9. We give an idea of the proof. We assume that

$$a = e \cdot p_1 \cdots p_r = e \cdot q_1 \cdots q_s$$

for $e \in \{\pm 1\}$ and for prime numbers $p_1, \ldots, p_r$, $q_1, \ldots, q_s$ ($r \in \mathbb{N}$, $s \in \mathbb{N}$), not necessarily distinct. Since now we have $p_1 \mid a$ and therefore $p_1 \mid e \cdot q_1 \cdots q_s$, it follows with the help of Euclid's lemma, Lemma 1.7, that $p_1 \mid q_j$ for some $j = 1, \ldots, s$. Since p_1 is prime, we must have $p_1 = q_j$. Without loss of generality, we may assume (by renumbering if necessary) that $p_1 = q_1$. Application of the cancellation law implies the equality

$$p_2 \cdots p_r = q_2 \cdots q_s. \tag{3}$$

Since p_2 divides the left-hand side of (3), p_2 must also divide the right-hand side. As in the first step, we conclude that $p_2 = q_2$. Proceeding in this way, we obtain the equalities $r = s$ and $p_j = q_j$ for $j = 1, \ldots, r$, which proves the asserted uniqueness.

Exercise 2.5. We leave it to the reader to prove that the polynomial ring $(R[X], +, \cdot)$ is a commutative ring if and only if $(R, +, \cdot)$ is commutative.

Exercise 2.6. Let A be a nonempty set and $(R, +_R, \cdot_R)$ a ring. Then, $0_R \in R$, and therefore, the mapping $0 : A \longrightarrow R$, $a \mapsto 0_R$, is an element of $\operatorname{map}(A, R)$. Hence, the set $\operatorname{map}(A, R)$ is not empty. We now show that $+$ on $\operatorname{map}(A, R)$ is associative. To this end, let $f, g, h \in \operatorname{map}(A, R)$. Then for all $a \in A$, we have

$$\begin{aligned}((f+g)+h)(a) &\underset{\text{def of } +}{=} (f+g)(a) +_R h(a) \underset{\text{def of } +}{=} (f(a) +_R g(a)) +_R h(a) \\ &\underset{\substack{+_R \text{ associative,} \\ \text{since } R \text{ is a ring}}}{=} f(a) +_R (g(a) +_R h(a)) \\ &\underset{\text{def of } +}{=} f(a) +_R (g+h)(a) \underset{\text{def of } +}{=} (f+(g+h))(a),\end{aligned}$$

which proves that $+$ is associative. The mapping $0 : A \longrightarrow R$ is the identity element with respect to $+$, since for all $f \in \operatorname{map}(A, R)$, we have the equality

$$(0+f)(a) \underset{\text{def. of } +}{=} 0(a) +_R f(a) \underset{\text{def. of } 0}{=} 0_R +_R f(a) \underset{\substack{0_R \text{ id. el.} \\ \text{w.r.t. } +_R}}{=} f(a)$$

and analogously the equality $(f+0)(a) = f(a)$ for all $a \in A$. The other ring properties of $\operatorname{map}(A, R)$ are proved similarly using the ring properties of R.

Exercise 2.7. The solution of this exercise is left to the reader.

Exercise 2.11. If $n > 1$ is not prime, then there exist natural numbers $a, b \in \mathcal{R}_n$, $a > 1$, $b > 1$, with $a \cdot b = n$. But then we have $a \odot b = 0$. Therefore, a and b are zero divisors of $\mathcal{R}_n$.

Exercise 2.12. We show here only that the lack of zero divisors in $(R, +, \cdot)$ implies the lack of zero divisors in $(R[X], +, \cdot)$. We assume that the ring $R[X]$ has zero divisors and show that then the ring R must have zero divisors. Let, then, $f(X) = a_n \cdot X^n + \cdots + a_1 \cdot X + a_0$ ($a_n \neq 0$) and $g(X) = b_m \cdot X^m + \cdots + b_1 \cdot X + b_0$ ($b_m \neq 0$) be zero divisors in $R[X]$, so that

$$f(X) \cdot g(X) = (a_n \cdot b_m) \cdot X^{n+m} + \cdots + (a_1 \cdot b_0 + a_0 \cdot b_1) \cdot X + a_0 \cdot b_0 = 0,$$

where 0 denotes the zero element of $R[X]$, i.e., the zero polynomial. In particular, we must have $a_n \cdot b_m = 0$, which proves that R has zero divisors.

Exercise 2.13. The ring $(\mathrm{map}(A, R), +, \cdot)$ from Exercise 2.6 has as its zero element $0 : A \longrightarrow R$, $a \mapsto 0_R$. If now, for example, we have $R = (\mathcal{R}_6, \oplus, \odot)$, then for $f : A \longrightarrow R$, $a \mapsto 2$ ($a \in A$) and $g : A \longrightarrow R$, $a \mapsto 3$ ($a \in A$), we have the equality

$$(f \cdot g)(a) \underset{\text{def of } \cdot}{=} f(a) \odot g(a) \underset{\text{def of } f,g}{=} 2 \odot 3 = 0 = 0(a),$$

that is, f and g are zero divisors of $(\mathrm{map}(A, R), +, \cdot)$. Therefore, the ring $(\mathrm{map}(A, R), +, \cdot)$ also has zero divisors if R has zero divisors. We note that the ring $(\mathrm{map}(A, R), +, \cdot)$ can possess zero divisors even if R has no zero divisors.

Exercise 2.19. In the polynomial ring $(\mathbb{Z}[X], +, \cdot)$, the unit element is given by the unit polynomial 1. Let $f(X) = a_n \cdot X^n + \cdots + a_1 \cdot X + a_0$ ($a_n \neq 0$) be a unit in $\mathbb{Z}[X]$. Then there exists a polynomial $g(X) = b_m \cdot X^m + \cdots + b_1 \cdot X + b_0$ ($b_m \neq 0$) in $\mathbb{Z}[X]$ such that

$$f(X) \cdot g(X) = (a_n \cdot b_m) \cdot X^{n+m} + \cdots + (a_1 \cdot b_0 + a_0 \cdot b_1) \cdot X + a_0 \cdot b_0 = 1.$$

If $n > 0$, then we must have in particular that $a_n \cdot b_m = 0$, in contradiction to the fact that $\mathbb{Z}$ has no zero divisors. Therefore, f and hence g must be of the form $f(X) = a_0$ and $g(X) = b_0$. The equality $a_0 \cdot b_0 = 1$ shows that f and g are units if and only if $a_0 \in \{1, -1\}$ and $a_0 = b_0$. The polynomial ring $(\mathbb{Z}[X], +, \cdot)$ has therefore only the units $\{1, -1\}$.

Exercise 2.20. That the units of a ring $(R, +, \cdot)$ with unit element 1 form a group under multiplication with multiplicative identity element 1 follows directly from how units are defined.

Exercise 2.21. The group of units of $\mathcal{R}_5$ is $(\mathcal{R}_5 \setminus \{0\}, \odot)$ and is therefore isomorphic to the group $(\mathcal{R}_4, \oplus)$. The group of units of $\mathcal{R}_8$ is $(\{1, 3, 5, 7\}, \odot)$.

We have $3 \odot 3 = 1$, $5 \odot 5 = 1$, $7 \odot 7 = 1$, which shows that this group is isomorphic to $(D_4, \circ)$. The group of units of $\mathcal{R}_{10}$ is $(\{1, 3, 7, 9\}, \odot)$. We have $3 \odot 7 = 1$, $9 \odot 9 = 1$, which shows that this group is isomorphic to $(\mathcal{R}_4, \oplus)$ and therefore to $(\mathcal{R}_5 \setminus \{0\}, \odot)$. The group of units of $\mathcal{R}_{12}$ is $(\{1, 5, 7, 11\}, \odot)$. We have $5 \odot 5 = 1$, $7 \odot 7 = 1$, $11 \odot 11 = 1$, which shows that this group is isomorphic to $(D_4, \circ)$ and therefore to $(\{1, 3, 5, 7\}, \odot)$.

Exercise 2.25. For $n \in \mathbb{N}$, we have that $(n\mathbb{Z}, +, \cdot)$ is a subring of $(\mathbb{Z}, +, \cdot)$.

Exercise 2.26. We leave it to the reader to show that $(R, +, \cdot)$ is a subring of the polynomial ring $(R[X], +, \cdot)$.

Exercise 3.2. (a) The mapping f_1 is a ring homomorphism.

(b) The mapping f_2 is not a ring homomorphism, since for $g_1(X) = X$ and $g_2(X) = 1$, we have $f_2(g_1(X) \cdot g_2(X)) = f_2(1 \cdot X) = 1 \neq 0 = 1 \cdot 0 = f_2(X) \cdot f_2(1) = f_2(g_1(X)) \cdot f_2(g_2(X))$.

(c) The mapping f_3 is a ring homomorphism if and only if $r = 0$.

(d) The mapping f_4 is a ring homomorphism.

(e) The mapping f_5 is a ring homomorphism.

Exercise 3.5. The proof of Lemma 3.4 is left to the reader.

Exercise 3.6. We obtain

$$\ker(f_1) = \Big\{ \sum_{j \in \mathbb{N}} a_j \cdot X^j \,\Big|\, a_j \in R, a_0 = 0 \Big\}, \quad \operatorname{im}(f_1) = R.$$

The mapping f_2 is not a ring homomorphism. Furthermore, we have

$$\begin{aligned}
&\ker(f_3) = \operatorname{map}(A, R), \quad \operatorname{im}(f_3) = \{0\} \text{ (if } r = 0); \\
&\ker(f_4) = \{g \in \operatorname{map}(A, R) \mid g(a) = 0\}, \quad \operatorname{im}(f_4) = \{g(a) \mid g \in \operatorname{map}(A, R)\}; \\
&\ker(f_5) = \{f(X) \in R[X] \mid f(r) = 0\}, \quad \operatorname{im}(f_5) = \{f(r) \mid f(X) \in R[X]\} = R,
\end{aligned}$$

for the ring homomorphisms under consideration.

Exercise 3.13. To prove the equality $\mathfrak{a} = R$, we must show that $\mathfrak{a} \supseteq R$. To this end, let $r \in R$. Since $1 \in \mathfrak{a}$ and $\mathfrak{a}$ is an ideal, it follows that $r \cdot 1 = r \in \mathfrak{a}$. This proves that $R \subseteq \mathfrak{a}$.

Exercise 3.14. No, since for every subring U of $(\mathbb{Z}, +, \cdot)$, we have in particular that $(U, +)$ is a subgroup of $(\mathbb{Z}, +)$. Therefore, we must have $U = n\mathbb{Z}$ for some $n \in \mathbb{N}$, which proves that $(U, +)$ is an ideal of $\mathbb{Z}$.

Exercise 3.15. For example, $\mathbb{Z}$ is a subring of the polynomial ring $(\mathbb{Z}[X], +, \cdot)$ that is not an ideal of $\mathbb{Z}[X]$.

Exercise 3.16. The principal ideals of $\mathbb{Z}[X]$ are of the form $\{h \cdot f \mid f \in \mathbb{Z}[X]\}$ for some $h \in \mathbb{Z}[X]$. We leave it to the reader to show that the ideal

$$\mathfrak{a} := \{2 \cdot f + X \cdot g \mid f, g \in \mathbb{Z}[X]\}$$

is not a principal ideal of $(\mathbb{Z}[X], +, \cdot)$.

Exercise 3.18. We have $\ker(f_1) = (X)$ and $\ker(f_5) = (X - r)$. If the ring R possesses a unit element, then $\ker(f_3) = (1)$.

Exercise 3.27. Let $a \in \mathbb{Z}$. Then the mapping $f : \mathbb{Z}[X] \longrightarrow \mathbb{Z}$ given by the assignment $f(X) \mapsto f(a)$ is a surjective ring homomorphism with $\ker(f) = (X - a)$ by Exercises 3.2, 3.6, and 3.18. Invoking Corollary 3.25, we see that $\mathbb{Z}[X]/(X - a) \cong \mathbb{Z}$ is a ring isomorphism, as desired.

Exercise 3.28. An analogy to the group isomorphism from Exercise 5.11 of Chapter II can also be formulated and proved for rings by replacing "subgroup" and "normal subgroup" throughout with "ideal" and observing that the group homomorphisms that arise are also ring homomorphisms. We leave this task to the reader.

Exercise 4.4. This is impossible, since one can show that every skew field with finitely many elements is in fact a field.

Exercise 5.3. The proofs of associativity of $\oplus$, commutativity of $\odot$, and the second distributive law are left to the reader.

Exercise 5.5. We consider the ring homomorphism $f : K \longrightarrow \mathrm{Quot}(K)$ given by $a \mapsto [a, 1]$. The ring homomorphism f is well defined, since K is a field, and therefore, we have $1 \in K$. Moreover, f is injective, which can be shown as follows: Assume that $[a, 1] = [b, 1]$ for $a, b \in K$. Then $(a, 1) \sim (b, 1)$, which implies $a \cdot 1 = 1 \cdot b$ and hence $a = b$. We now show that f is also surjective. To this end, let $[a, b] \in \mathrm{Quot}(K)$ with $a, b \in K$, $b \neq 0$, be an arbitrary element. Since K is a field, there exists an inverse $b^{-1} \in K$ to b. Therefore, we have $a \cdot b^{-1} \in K$ and

$$f(a \cdot b^{-1}) = [a \cdot b^{-1}, 1] = [a, b],$$

where the last equality follows from $(a \cdot b^{-1}, 1) \sim (a, b) \iff a \cdot b^{-1} \cdot b = 1 \cdot a \iff a = a$. Therefore, f is a ring isomorphism, and we have the ring isomorphism $K \cong \mathrm{Quot}(K)$.

Exercise 6.3. Let $r = s/t$ for $s \in \mathbb{Z}$ and $t \in \mathbb{N} \setminus \{0\}$. Then r corresponds to the element $[s, t] \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$. If now $d = (s, t) > 0$ is the greatest common divisor of s and t, then we may write $s = d \cdot a$ and $t = d \cdot b$ with $a \in \mathbb{Z}$, $b \in \mathbb{N} \setminus \{0\}$, and a, b relatively prime. From the equality $s \cdot b = (d \cdot a) \cdot b = (d \cdot b) \cdot a = t \cdot a$ we infer the equality $[s, t] = [a, b]$. This proves the existence

of the claimed representative. To prove the uniqueness of the representative, we assume that $[c, d]$ with $c \in \mathbb{Z}$, $d \in \mathbb{N} \setminus \{0\}$, and c, d relatively prime is also an element such that $[s, t] = [c, d]$. Then in particular, we have $[a, b] = [c, d]$, and therefore

$$a \cdot d = b \cdot c.$$

If p is a prime number such that $p \mid a$, then we must have $p \mid b \cdot c$ and consequently, since p is prime, $p \mid b$ or $p \mid c$. But since a and b are relatively prime, we must have $p \mid c$. Conversely, one can show that for an arbitrary prime p with $p \mid c$, we must also have $p \mid a$. From this it follows that $a \mid c$ and also $c \mid a$, which, since a and c have the same sign, proves the equality $a = c$. We conclude by an analogous argument the equality $b = d$, completing the proof of uniqueness.

Exercise 6.4. Since one usually learns this proof in a first course in real analysis, we leave this exercise to the reader.

Exercise 6.7. The generalizations of the addition and multiplication rules from Remark 1.19 of Chapter I are left to the reader.

Exercise 7.4. We obtain in $\mathbb{Z}[X]$ the decompositions $20X = 2^2 \cdot 5 \cdot X$ and $10X^2 + 4X - 6 = 2 \cdot (X + 1) \cdot (5X - 3)$. Therefore, 2 is the greatest common divisor, and $2^2 \cdot 5 \cdot X \cdot (X + 1) \cdot (5X - 3) = 100X^3 + 40X^2 - 60X$ the least common multiple, of the polynomials $20X$ and $10X^2 + 4X - 6$ in $\mathbb{Z}[X]$.

Exercise 7.9. We leave it to the reader to come up with relevant examples.

Exercise 7.15. Part (ii) of the proof of Lemma 7.14 is left to the reader.

Exercise 7.25. For example, the polynomial ring $K[X, Y]$ in two variables over a field K is not a principal ideal domain, since the ideal

$$\mathfrak{a} := \{X \cdot f + Y \cdot g \mid f, g \in K[X, Y]\}$$

is not principal.

Exercise 7.38. (a) We obtain $(123456789, 555555555) = 9$.
(b) We calculate

$$X^4 + 2X^3 + 2X^2 + 2X + 1 = (X + 1) \cdot (X^3 + X^2 - X - 1) + (2X^2 + 4X + 2),$$
$$X^3 + X^2 - X - 1 = \left(\frac{1}{2}X - \frac{1}{2}\right) \cdot (2X^2 + 4X + 2) + 0.$$

We thereby obtain $(X^4 + 2X^3 + 2X^2 + 2X + 1, X^3 + X^2 - X - 1) = 2X^2 + 4X + 2$.

Solutions to Exercises in Chapter IV

Exercise 1.6. (a) We obtain the decimal expansions $\frac{1}{5} = 0.2$, $\frac{1}{3} = 0.\overline{3}$, $\frac{1}{16} = 0.0625$, $\frac{1}{11} = 0.\overline{09}$, and $\frac{1}{7} = 0.\overline{142857}$.

(b) One shows that a reduced fraction $\frac{a}{b}$ ($a, b \in \mathbb{Z}; b \neq 0$) has a terminating decimal expansion if and only if $b = 2^k \cdot 5^l$ with $k, l \in \mathbb{N}$.

(c) We consider the fraction $\frac{1}{m}$ for $m \in \mathbb{N}$, $m \neq 0$, with $2 \nmid m$ and $5 \nmid m$. One shows that $m - 1$ is the maximal period length of the decimal expansion of the fraction $\frac{1}{m}$, considering that there can be at most $m - 1$ remainders on division by m. The fraction $\frac{1}{7}$, for example, has a period length that is maximal.

(d) Without loss of generality, we may assume that the periodic decimal fraction has the form

$$0.q_{-1} \dots q_{-v}\overline{q_{-(v+1)} \cdots q_{-(v+p)}}$$

with natural numbers $v \geq 0$, $p > 0$. Then

$$\frac{a}{b} = \frac{\sum_{j=1}^{v} q_{-j} 10^{v-j}}{10^v} + \frac{1}{10^v} \cdot \frac{\sum_{j=1}^{p} q_{-(v+j)} 10^{p-j}}{10^p - 1}.$$

For example, for $0.\overline{123}$, one obtains the fraction

$$\frac{123}{10^3 - 1} = \frac{123}{999} = \frac{41}{333}.$$

Exercise 2.2. (a) Without loss of generality, we may assume that $\epsilon \in \mathbb{Q}$, $0 < \epsilon < 1$. For $n \in \mathbb{N}$, we then have

$$\left| \frac{1}{n+1} \right| < \epsilon \Longleftrightarrow \frac{1-\epsilon}{\epsilon} < n.$$

If we set $N(\epsilon) := [(1 - \epsilon)/\epsilon]$, where $[x]$ is the greatest integer less than or equal to x, then for all $n \in \mathbb{N}$ with $n > N(\epsilon)$, we have the inequality

$$\left| \frac{1}{n+1} \right| < \epsilon.$$

This proves that the sequence $\left(\frac{1}{n+1}\right)_{n \geq 0}$ is a rational null sequence. Using the inequality

$$\frac{n}{2^n} < \frac{1}{n+1}$$

for $n \in \mathbb{N}$, $n \geq 5$, one shows analogously that $\left(\frac{n}{2^n}\right)_{n \geq 0}$ is also a rational null sequence.

(b) Other examples of rational null sequences are the sequences $\left(\frac{1}{(n+1)^k}\right)_{n\geq 0}$ with $k \in \mathbb{N}$, $k \geq 2$.

Exercise 2.18. Let $\alpha = (a_n) + \mathfrak{n}$, $\beta = (b_n) + \mathfrak{n}$ be two real numbers with $\alpha < \beta$, that is, there exist $q \in \mathbb{Q}$, $q > 0$, $N(q) \in \mathbb{N}$ with $b_n - a_n > q$ for all $n \in \mathbb{N}$ with $n > N(q)$. To prove the asserted independence of the choice of the representing rational Cauchy sequences, we assume that $(a_n) + \mathfrak{n} = (a'_n) + \mathfrak{n}$ and $(b_n) + \mathfrak{n} = (b'_n) + \mathfrak{n}$ for rational Cauchy sequences (a'_n) and (b'_n). Then in particular, we must have $(a'_n) = (a_n) + (c_n) = (a_n + c_n)$ and $(b'_n) = (b_n) + (d_n) = (b_n + d_n)$ for rational null sequences (c_n) and (d_n). This yields

$$b'_n - a'_n = (b_n - a_n) + (d_n - c_n) \geq (b_n - a_n) - |d_n - c_n|. \tag{4}$$

Since, moreover, $(d_n - c_n)$ is a rational null sequence, there exists for $\epsilon := q/2 \in \mathbb{Q}$ an $\widetilde{N}(\epsilon) \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n > \widetilde{N}(\epsilon)$, we have the inequality $|d_n - c_n| < \epsilon$. So if we set $q' := q - \epsilon = q/2 \in \mathbb{Q}$, we have $q' > 0$, and with (4) we obtain the inequality

$$b'_n - a'_n > q - \epsilon = q'$$

for all $n \in \mathbb{N}$ with $n > N(q') := \max(N(q), \widetilde{N}(\epsilon))$. This proves the claimed independence of representative.

Exercise 2.22. We leave the proof of Lemma 2.21 to the reader.

Exercise 2.25. Real null sequences whose elements are irrational numbers include, for example, the sequences $\left(\frac{\sqrt{2}}{(n+1)^k}\right)_{n\geq 0}$ with $k \in \mathbb{N}$, $k \geq 1$.

Exercise 2.30. We begin by considering that for a rational number $a_0 \in \mathbb{Q}$, $a_0 > 0$, we have the following equivalences with an error of δ_0:

$$\begin{aligned} a_0 + \delta_0 = \sqrt{2} &\iff (a_0 + \delta_0)^2 = 2 \iff 2a_0\delta_0 + \delta_0^2 = 2 - a_0^2 \\ &\iff \delta_0 = \frac{2 - a_0^2}{2a_0} - \frac{\delta_0^2}{2a_0}. \end{aligned}$$

So if we set

$$a_1 := a_0 + \frac{2 - a_0^2}{2a_0} = \frac{2 + a_0^2}{2a_0},$$

then because $a_0 > 0$, we have the inequality $a_1^2 > 2$, that is, $a_1 > \sqrt{2}$. This implies $\left(2 - a_1^2\right)/(2a_1) < 0$, and we therefore have for

$$a_2 := a_1 + \frac{2 - a_1^2}{2a_1} = \frac{2 + a_1^2}{2a_1}$$

both $a_1 > a_2$ and $a_2^2 > 2$, that is, $a_2 > \sqrt{2}$. We consider now the rational sequence $(a_n)_{n\geq 0}$ with $a_0 \in \mathbb{Q}$, $a_0 > 0$, arbitrary and

$$a_{n+1} := \frac{2 + a_n^2}{2a_n} \quad (n \in \mathbb{N}, n \geq 1). \tag{5}$$

One first shows by induction that both $a_n > a_{n+1}$ for all $n \in \mathbb{N}$, $n \geq 1$, and $a_n^2 > 2$, that is, $a_n > \sqrt{2}$, for all $n \in \mathbb{N}$, $n \geq 1$. Using these inequalities, one then shows in a second step that $(a_n)_{n\geq 0}$ is a rational Cauchy sequence. This rational Cauchy sequence has limit $\alpha \in \mathbb{R}$, $\alpha > 0$. Because of the recurrence formula (5), we see that α satisfies the equation

$$\alpha = \frac{2 + \alpha^2}{2\alpha} \Longleftrightarrow \alpha = \sqrt{2},$$

as desired.

Exercise 3.12. (a) The decimal representation $0.101\,001\,000\,100\,001\ldots$ is neither terminating nor periodic. Therefore, this number cannot be rational. One can find analogous examples, such as $0.121\,331\,222\,133\,331\ldots$.

(b) Using the rational Cauchy series $(a_n)_{n\geq 0}$ constructed in Exercise 2.30 for calculating $\sqrt{2}$, one obtains $\sqrt{2} \approx 1.4142135623$, accurate to ten decimal places, by choosing, for example, $a_0 = 1$ and iterating four times.

Exercise 4.2. We consider, for example, the sequence

$$(a_n)_{n\geq 0} := \left(\frac{n^2 + 2}{2^n}\right)_{n\geq 0}.$$

We have $a_0 = 2$ and $a_1 = a_2 = 3/2$. For $n \in \mathbb{N}$, $n \geq 3$, one shows the inequality $2(n^2 + 2) > (n + 1)^2 + 2$. Since

$$2(n^2 + 2) > (n + 1)^2 + 2 \Longleftrightarrow \frac{n^2 + 2}{2^n} > \frac{(n + 1)^2 + 2}{2^{n+1}} \Longleftrightarrow a_n > a_{n+1}$$

for $n \in \mathbb{N}$, $n \geq 3$, it follows that the sequence $(a_n)_{n\geq 0}$ is monotonically decreasing, but not strictly. One shows analogously that the sequences

$$\left(12^{\frac{1}{n+1}}\right)_{n\geq 0}, \left(\frac{n^3 + 3}{3^n}\right)_{n\geq 0}$$

are strictly monotonically decreasing and that the sequence

$$\left(\frac{n^3 - 2}{n^2 - 2}\right)_{n\geq 0}$$

is monotonically increasing. The sequence

$$\left(n^{\frac{1}{n+1}}\right)_{n\geq 0}$$

is neither monotonically increasing nor monotonically decreasing. However, the sequence

$$\left(n^{\frac{1}{n+1}}\right)_{n\geq 4}$$

is strictly monotonically decreasing.

Exercise 4.6. The subset $\mathfrak{M} \subseteq \mathbb{Q}$ consisting of all sequence terms a_n $(n > 0)$ in the rational Cauchy sequence $(a_n)_{n\geq 0}$ constructed in Exercise 2.30, which is bounded below, has greatest lower bound $\sqrt{2} \notin \mathbb{Q}$.

Exercise 4.7. The greatest lower bound of the set $\{\sqrt[x]{x} \mid x \in \mathbb{Q}, x \geq 0\}$ is attained when $x = 0$ and is equal to zero. The least upper bound is $\sqrt[e]{e}$.

Exercise 5.3. The solution of this exercise is left to the reader.

Exercise 6.7. We leave the completion of the details in the sketch of the proof of Theorem 6.5 to the reader.

Solutions to Exercises in Chapter V

Exercise 1.1. The solution of this exercise is left to the reader.

Exercise 1.8. The completion of the proof of Theorem 1.7 is left to the reader.

Exercise 1.10. Let $\alpha = \alpha_1 + \alpha_2 i \in \mathbb{C}$, $\alpha \neq 0$. If $\alpha_2 = 0$ and $\alpha_1 > 0$, then $\pm\sqrt{\alpha_1}$ are the solutions to the equation $x^2 = \alpha$. If $\alpha_2 = 0$ and $\alpha_1 < 0$, then $\pm\sqrt{|\alpha_1|}i$ are the solutions to the equation $x^2 = \alpha$. It remains to consider the case $\alpha_2 \neq 0$. Let $\beta = \beta_1 + \beta_2 i \in \mathbb{C}$ with $\beta_1 \neq 0$. We then have the equivalence

$$\beta^2 = \alpha \iff (\beta_1^2 - \beta_2^2) + (2\beta_1\beta_2)i = \alpha_1 + \alpha_2 i \iff \beta_1^2 - \beta_2^2 = \alpha_1,\ 2\beta_1\beta_2 = \alpha_2.$$

If we substitute the second equation, $\beta_2 = \alpha_2/(2\beta_1)$, in the first equation, we obtain the equation $4\beta_1^4 - 4\alpha_1\beta_1^2 - \alpha_2^2 = 0$. Setting $y := \beta_1^2$, we obtain the quadratic equation $4y^2 - 4\alpha_1 y - \alpha_2^2 = 0$, which has the solutions

$$y_{1,2} = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + \alpha_2^2}}{2} = \frac{\alpha_1 \pm |\alpha|}{2}.$$

Since now $\beta_1 \in \mathbb{R}$, we need consider only the nonnegative solution y_1. With $\beta_1 = \pm\sqrt{y_1}$ and $\beta_2 = \pm\alpha_2/(2\sqrt{y_1})$, we obtain the solution formula

$$\beta = \pm\frac{\sqrt{\alpha_1 + |\alpha|}}{\sqrt{2}} \pm \frac{\alpha_2 i}{\sqrt{2\big(\alpha_1 + |\alpha|\big)}}.$$

Altogether, we obtain for the solutions to the equation $x^2 = \alpha$ the following solution formula:

$$x_{1,2} = \begin{cases} \pm\sqrt{\alpha_1}, & \text{if } \alpha_1 > 0, \alpha_2 = 0; \\ \pm\sqrt{|\alpha_1|}\, i, & \text{if } \alpha_1 < 0, \alpha_2 = 0; \\ \pm\Big(\sqrt{\frac{|\alpha|+\alpha_1}{2}} + \sqrt{\frac{|\alpha|-\alpha_1}{2}}\, i\Big), & \text{if } \alpha_2 > 0; \\ \pm\Big(\sqrt{\frac{|\alpha|+\alpha_1}{2}} - \sqrt{\frac{|\alpha|-\alpha_1}{2}}\, i\Big), & \text{if } \alpha_2 < 0. \end{cases}$$

From this it follows that the solutions to the equation $x^2 = i$ are

$$x_{1,2} = \pm\frac{1+i}{\sqrt{2}},$$

those to the equation $x^2 = 2 + i$ are

$$x_{1,2} = \pm\Big(\frac{\sqrt{\sqrt{5}+2}}{\sqrt{2}} + \frac{\sqrt{\sqrt{5}-2}}{\sqrt{2}}\, i\Big),$$

and those to the equation $x^2 = 3 - 2i$ are

$$x_{1,2} = \pm\Big(\frac{\sqrt{\sqrt{13}+3}}{\sqrt{2}} - \frac{\sqrt{\sqrt{13}-3}}{\sqrt{2}}\, i\Big).$$

Exercise 1.11. Since by Exercise 1.10 we have the equality $((1+i)/2)^2 = i/2$, we obtain, on completing the square,

$$x^2 + (1+i)\cdot x + i = 0 \Longleftrightarrow \Big(x + \frac{1+i}{2}\Big)^2 + \frac{i}{2} = 0.$$

If we now substitute $y := x + (1+i)/2$, we obtain the quadratic equation $y^2 = -i/2$. With the solution formula from Exercise 1.10, we obtain the solution

$$x_{1,2} = y_{1,2} - \frac{1+i}{2} = \pm\frac{1}{2} \mp \frac{i}{2} - \frac{1+i}{2},$$

that is, $x_1 = -i$ and $x_2 = -1$ are the solutions of $x^2 + (1+i)\cdot x + i = 0$, which can be easily checked by substitution. One can prove analogously that the equation $x^2 + (2-i)\cdot x - 2i = 0$ has the solutions $x_1 = i$ and $x_2 = -2$.

Exercise 1.14. (a) First, one verifies the equality $\overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta}$ for all $\alpha, \beta \in \mathbb{C}$. One then has

$$|\alpha \cdot \beta|^2 = (\alpha \cdot \beta) \cdot \overline{\alpha \cdot \beta} = \alpha \cdot \overline{\alpha} \cdot \beta \cdot \overline{\beta} = |\alpha|^2 \cdot |\beta|^2,$$

which proves the assertion.
(b) Let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{N}$. The product rule from part (a) with $\alpha = \alpha_1 + \alpha_2 i$ and $\beta = \beta_1 + \beta_2 i$ yields

$$(\alpha_1^2 + \alpha_2^2) \cdot (\beta_1^2 + \beta_2^2) = (\alpha_1\beta_1 - \alpha_2\beta_2)^2 + (\alpha_1\beta_2 + \alpha_2\beta_1)^2.$$

This implies the assertion.

Exercise 2.3. The statement of Remark 2.2 results immediately from use of the product rule from Exercise 1.14.

Exercise 2.5. Let $A \in \mathrm{M}_2(\mathbb{R})$ be an invertible matrix. One may convince oneself that the mapping $f_A : (\mathbb{C}, +, \cdot) \longrightarrow (\mathrm{M}_2(\mathbb{R}), +, \cdot)$ given by

$$\alpha = \alpha_1 + \alpha_2 i \mapsto A \cdot \begin{pmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{pmatrix} \cdot A^{-1}$$

is an injective ring homomorphism. In particular, f induces an isomorphism $\mathbb{C} \cong \mathrm{im}(f)$, that is, $\mathbb{C}$ is isomorphic to the subring $\mathrm{im}(f)$ of $\mathrm{M}_2(\mathbb{R})$.

Exercise 2.8. The solution of this exercise is left to the reader.

Exercise 2.12. We first observe that for two complex numbers $\alpha, \beta \in \mathbb{C} \setminus \{0\}$ with polar coordinate representations $\alpha = |\alpha| \cdot (\cos(\varphi) + i\sin(\varphi))$ and $\beta = |\beta| \cdot (\cos(\psi) + i\sin(\psi))$, one has the following multiplication formula:

$$\begin{aligned} \alpha \cdot \beta &= |\alpha||\beta| \cdot \big(\cos(\varphi)\cos(\psi) - \sin(\varphi)\sin(\psi) \\ &\quad + i\big(\sin(\varphi)\cos(\psi) + \cos(\varphi)\sin(\psi)\big)\big) \\ &= |\alpha||\beta| \cdot \big(\cos(\varphi + \psi) + i\sin(\varphi + \psi)\big), \end{aligned}$$

where for the second equality we have invoked the addition theorems for sine and cosine. From this follows for $m \in \mathbb{N}$ the equality

$$\alpha^m = |\alpha|^m \cdot \big(\cos(m\varphi) + i\sin(m\varphi)\big).$$

Likewise, for $n \in \mathbb{N}$ with $n \neq 0$, we obtain the equality

$$\alpha^{\frac{1}{n}} = |\alpha|^{\frac{1}{n}} \cdot \Big(\cos\Big(\frac{\varphi}{n}\Big) + i\sin\Big(\frac{\varphi}{n}\Big)\Big).$$

This completes the proof of the general formula.

Exercise 4.4. Let p be a prime number. Then $f(X) = X^2 - p$ is a quadratic polynomial with integer coefficients, and we have $f(\sqrt{p}) = 0$. We now assume that there exists a linear polynomial $g(X) = aX + b$ ($a, b \in \mathbb{Z}$, $a \neq 0$) with $g(\sqrt{p}) = 0$. But then we must have $\sqrt{p} = -b/a \in \mathbb{Q}$, a contradiction. We have therefore shown that $\sqrt{p}$ is algebraic of degree 2.

Exercise 4.12. We leave it to the reader to find additional transcendental numbers following the pattern of the Liouville number.

Exercise 5.8. The calculation of better approximations to e is left to the reader.

Solutions to Exercises in Chapter VI

Exercise 1.6. For example, the quadratic polynomial $X^2 + 1 \in \mathbb{H}[X]$ has zeros $\pm i, \pm j, \pm k$ and every purely imaginary quaternion $\alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathrm{Im}(\mathbb{H})$, that satisfies the condition $\alpha_2^2 + \alpha_3^2 + \alpha_4^2 = 1$.

Exercise 1.7. It is clear that $\mathbb{R} \subseteq Z(\mathbb{H})$. We therefore have to show that $Z(\mathbb{H}) \subseteq \mathbb{R}$. To this end, let $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in Z(\mathbb{H})$. For each $\beta = \beta_1 + \beta_2 i + \beta_3 j + \beta_4 k \in \mathbb{H}$, we then have $\alpha \cdot \beta = \beta \cdot \alpha$. Since

$$\begin{aligned}\alpha \cdot \beta = {} & (\alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3 - \alpha_4\beta_4) + (\alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_4 - \alpha_4\beta_3)i \\ & + (\alpha_1\beta_3 - \alpha_2\beta_4 + \alpha_3\beta_1 + \alpha_4\beta_2)j + (\alpha_1\beta_4 + \alpha_2\beta_3 - \alpha_3\beta_2 + \alpha_4\beta_1)k\end{aligned}$$

and

$$\begin{aligned}\beta \cdot \alpha = {} & (\beta_1\alpha_1 - \beta_2\alpha_2 - \beta_3\alpha_3 - \beta_4\alpha_4) + (\beta_1\alpha_2 + \beta_2\alpha_1 + \beta_3\alpha_4 - \beta_4\alpha_3)i \\ & + (\beta_1\alpha_3 - \beta_2\alpha_4 + \beta_3\alpha_1 + \beta_4\alpha_2)j + (\beta_1\alpha_4 + \beta_2\alpha_3 - \beta_3\alpha_2 + \beta_4\alpha_1)k,\end{aligned}$$

however, we have

$$\begin{aligned}\alpha \cdot \beta = \beta \cdot \alpha &\iff 2(\alpha_3\beta_4 - \alpha_4\beta_3)i + 2(-\alpha_2\beta_4 + \alpha_4\beta_2)j + 2(\alpha_2\beta_3 - \alpha_3\beta_2)k = 0 \\ &\iff \alpha_3\beta_4 = \alpha_4\beta_3 \wedge \alpha_4\beta_2 = \alpha_2\beta_4 \wedge \alpha_2\beta_3 = \alpha_3\beta_2.\end{aligned}$$

If $\alpha_2 \neq 0$, then from the third equality, it follows that $\beta_3 = (\alpha_3\alpha_2^{-1}) \cdot \beta_2$ for every $\beta \in \mathbb{H}$. This contradiction implies that we must have $\alpha_2 = 0$. Similarly, one shows that we must also have $\alpha_3 = 0$ and $\alpha_4 = 0$. Altogether, therefore, we have that $\alpha = \alpha_1 \in \mathbb{R}$. This proves the inclusion $Z(\mathbb{H}) \subseteq \mathbb{R}$.

Exercise 1.14. (a) Let $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k$. We then calculate

$$\begin{aligned}\alpha^2 &= (\alpha_1^2 - \alpha_2^2 - \alpha_3^2 - \alpha_4^2) + 2\alpha_1\alpha_2 i + 2\alpha_1\alpha_3 j + 2\alpha_1\alpha_4 k \\ &= -(\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2) + 2\alpha_1\alpha,\end{aligned}$$

which yields the desired result.
(b) This can be established by a direct calculation.

Exercise 1.15. Let $\alpha = \mathrm{Im}(\alpha) \cdot \mathrm{i}$ with $\mathrm{Im}(\alpha) = (\alpha_2, \alpha_3, \alpha_4)$ and $\beta = \mathrm{Im}(\beta) \cdot \mathrm{i}$ with $\mathrm{Im}(\beta) = (\beta_2, \beta_3, \beta_4)$. We calculate

$$\begin{aligned}\alpha \cdot \beta &= (\alpha_2 i + \alpha_3 j + \alpha_4 k) \cdot (\beta_2 i + \beta_3 j + \beta_4 k) \\ &= (-\alpha_2\beta_2 - \alpha_3\beta_3 - \alpha_4\beta_4) + (\alpha_3\beta_4 - \alpha_4\beta_3) i \\ &\quad + (-\alpha_2\beta_4 + \alpha_4\beta_2) j + (\alpha_2\beta_3 - \alpha_3\beta_2) k \\ &= -\left\langle \mathrm{Im}(\alpha)^t, \mathrm{Im}(\beta)^t \right\rangle + \left(\mathrm{Im}(\alpha)^t \times \mathrm{Im}(\beta)^t\right) \cdot \mathrm{i},\end{aligned}$$

which proves the assertion.

Exercise 1.18. Let $\alpha, \beta \in \mathbb{H}$. We calculate

$$2 \cdot \langle \overline{\alpha}, \beta \rangle = 2 \cdot \mathrm{Re}(\overline{\alpha} \cdot \overline{\beta}) = \overline{\alpha} \cdot \overline{\beta} + \overline{\overline{\alpha} \cdot \overline{\beta}} = \overline{\alpha} \cdot \overline{\beta} + \beta \cdot \alpha.$$

Multiplication of this equality on the right by β yields

$$2 \cdot \langle \overline{\alpha}, \beta \rangle \cdot \beta = \overline{\alpha} \cdot \overline{\beta} \cdot \beta + \beta \cdot \alpha \cdot \beta,$$

which on account of $\overline{\beta} \cdot \beta = \langle \beta, \beta \rangle \in \mathbb{R}$ proves the result.

Exercise 1.19. (a) The equality $\overline{\alpha \cdot \beta} = \overline{\beta} \cdot \overline{\alpha}$ can be verified by a direct calculation.
(b) Using part (a), one obtains

$$|\alpha \cdot \beta|^2 = (\alpha \cdot \beta) \cdot \overline{\alpha \cdot \beta} = \alpha \cdot (\beta \cdot \overline{\beta}) \cdot \overline{\alpha} = |\beta|^2 \cdot (\alpha \cdot \overline{\alpha}) = |\beta|^2 \cdot |\alpha|^2 = |\alpha|^2 \cdot |\beta|^2,$$

which proves the assertion.
(c) Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2, \beta_3, \beta_4 \in \mathbb{N}$. Using the product rule from part (a), we obtain

$$\begin{aligned}&(\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2) \cdot (\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2) \\ &\quad = (\alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3 - \alpha_4\beta_4)^2 + (\alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_4 - \alpha_4\beta_3)^2 \\ &\qquad + (\alpha_1\beta_3 + \alpha_3\beta_1 + \alpha_4\beta_2 - \alpha_2\beta_4)^2 + (\alpha_1\beta_4 + \alpha_4\beta_1 + \alpha_2\beta_3 - \alpha_3\beta_2)^2.\end{aligned}$$

This implies the assertion.

Exercise 1.21. The completion of the proof of Theorem 1.20 is left to the reader.

Exercise 1.25. Verification of the assertions of this problem is left to the reader.

Exercise 1.27. The solution to this problem is left to the reader.

Exercise 2.3. The statement of Remark 2.2 is an immediate result of the product rule from Exercise 1.19.

Exercise 2.5. Let $A \in \mathrm{M}_2(\mathbb{C})$ be an arbitrary invertible matrix. One may convince oneself that the mapping $f : (\mathbb{H}, +, \cdot) \longrightarrow (\mathrm{M}_2(\mathbb{C}), +, \cdot)$ given by

$$\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \mapsto A \cdot \begin{pmatrix} \alpha_1 + \alpha_2 i & \alpha_3 + \alpha_4 i \\ -\alpha_3 + \alpha_4 i & \alpha_1 - \alpha_2 i \end{pmatrix} \cdot A^{-1}$$

is an injective ring homomorphism. In particular, f induces an isomorphism $\mathbb{H} \cong \mathrm{im}(f)$, that is, $\mathbb{H}$ is isomorphic to the subring $\mathrm{im}(f)$ of $\mathrm{M}_2(\mathbb{C})$.

Exercise 2.6. We show that f is an $\mathbb{R}$-linear mapping. Let $\alpha = \alpha_1 + \alpha_2 i + \alpha_3 j + \alpha_4 k \in \mathbb{H}$ and $\beta = \beta_1 + \beta_2 i + \beta_3 j + \beta_4 k \in \mathbb{H}$. For arbitrary $\mu, \nu \in \mathbb{R}$, one has

$$\begin{aligned} f(\mu\alpha + \nu\beta) &= f((\mu\alpha_1 + \nu\beta_1) + (\mu\alpha_2 + \nu\beta_2)i + (\mu\alpha_3 + \nu\beta_3)j + (\mu\alpha_4 + \nu\beta_4)k) \\ &= \begin{pmatrix} (\mu\alpha_1 + \nu\beta_1) + (\mu\alpha_2 + \nu\beta_2)i & (\mu\alpha_3 + \nu\beta_3) + (\mu\alpha_4 + \nu\beta_4)i \\ -(\mu\alpha_3 + \nu\beta_3) + (\mu\alpha_4 + \nu\beta_4)i & (\mu\alpha_1 + \nu\beta_1) - (\mu\alpha_2 + \nu\beta_2)i \end{pmatrix} \\ &= \mu \begin{pmatrix} \alpha_1 + \alpha_2 i & \alpha_3 + \alpha_4 i \\ -\alpha_3 + \alpha_4 i & \alpha_1 - \alpha_2 i \end{pmatrix} + \nu \begin{pmatrix} \beta_1 + \beta_2 i & \beta_3 + \beta_4 i \\ -\beta_3 + \beta_4 i & \beta_1 - \beta_2 i \end{pmatrix} \\ &= \mu f(\alpha) + \nu f(\beta), \end{aligned}$$

that is, f is $\mathbb{R}$-linear. The verification of the remaining assertions are left to the reader.

Exercise 2.9. The solution to this problem is left to the reader.

Exercise 3.4. The solution to this problem is left to the reader.

Exercise 3.6. We first prove that every $\mathbb{R}$-linear mapping $v \mapsto A \cdot v$ ($v \in \mathbb{R}^3$) with $A \in \mathrm{SO}_3(\mathbb{R})$ is an orientation-preserving rotation of $\mathbb{R}^3$ about an axis passing through the origin. To this end, we begin by noting that for $A \in \mathrm{SO}_3(\mathbb{R})$, on account of

$$\begin{aligned} \det(A - E) &= 1 \cdot \det(A - E) = \det(A^t) \cdot \det(A - E) \\ &= \det(A^t \cdot (A - E)) = \det(E - A^t) = \det(E - A)^t \\ &= \det(E - A) = (-1) \cdot \det(A - E), \end{aligned}$$

we have the equality $\det(A - E) = 0$, which proves that 1 is an eigenvalue of A. Now let a_1 denote a normalized (to have length 1) eigenvector of A associated with the eigenvalue 1. We shall see that the mapping $v \mapsto A \cdot v$ ($v \in \mathbb{R}^3$) describes a rotation about an axis passing through the origin that is determined by a_1. For this, we extend a_1 to an orthonormal basis of $\mathbb{R}^3$ by choosing an additional vector $a_2 \in \mathbb{R}^3$, normed to have length 1, that is perpendicular to a_1, and then setting $a_3 := a_1 \times a_2$. If now $S \in \mathrm{M}_2(\mathbb{R})$ denotes a matrix with $S \cdot (1,0,0)^t = a_1$, $S \cdot (0,1,0)^t = a_2$, and $S \cdot (0,0,1)^t = a_3$, then we must have $S \in \mathrm{SO}_3(\mathbb{R})$. We further obtain

$$S^{-1} \cdot A \cdot S \cdot (1,0,0)^t = S^{-1} \cdot A \cdot a_1 = S^{-1} \cdot a_1 = (1,0,0)^t,$$

which implies the equality

$$S^{-1} \cdot A \cdot S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & \gamma & \delta \end{pmatrix}$$

for certain $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. But since because of $A, S \in \mathrm{SO}_3(\mathbb{R})$, we have also the inclusion $S^{-1} \cdot A \cdot S \in \mathrm{SO}_3(\mathbb{R})$ and thereby

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SO}_2(\mathbb{R}),$$

it follows that there exists a uniquely determined $\varphi \in [0, 2\pi)$ with

$$S^{-1} \cdot A \cdot S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\varphi) & -\sin(\varphi) \\ 0 & \sin(\varphi) & \cos(\varphi) \end{pmatrix} =: D_\varphi.$$

The mapping $v \mapsto D_\varphi \cdot v$ ($v \in \mathbb{R}^3$) is an orientation-preserving rotation in the x_2, x_3-plane through the angle φ about the x_1-axis (counterclockwise if a_1 points toward the observer). Altogether, we have shown that the mapping $v \mapsto A \cdot v$ ($v \in \mathbb{R}^3$) is an orientation-preserving rotation in the a_2, a_3-plane through the angle φ about the a_1-axis (counterclockwise if a_1 points toward the observer).

We now describe, conversely, an arbitrary orientation-preserving rotation of $\mathbb{R}^3$ through the angle $\varphi \in [0, 2\pi)$ about an axis passing through the origin that is determined by the vector (normed to have length 1) $a_1 := (\nu_1, \nu_2, \nu_3)^t \in \mathbb{R}^3$. To this end, we first consider the matrices

$$D_1 := \begin{pmatrix} \frac{\nu_1}{\sqrt{\nu_1^2+\nu_3^2}} & 0 & \frac{\nu_3}{\sqrt{\nu_1^2+\nu_3^2}} \\ 0 & 1 & 0 \\ -\frac{\nu_3}{\sqrt{\nu_1^2+\nu_3^2}} & 0 & \frac{\nu_1}{\sqrt{\nu_1^2+\nu_3^2}} \end{pmatrix}, \quad D_2 := \begin{pmatrix} \sqrt{\nu_1^2+\nu_3^2} & \nu_2 & 0 \\ -\nu_2 & \sqrt{\nu_1^2+\nu_3^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{SO}_3(\mathbb{R}).$$

If in the first step we apply D_1 to a_1, we rotate a_1 about the x_2-axis in such a way that $D_1 \cdot a_1$ lies in the x_1, x_2-plane. If in the second step we apply D_2 to $D_1 \cdot a_1$, then we rotate $D_1 \cdot a_1$ about the x_3-axis, so that finally, $D_2 \cdot D_1 \cdot a_1$ is parallel to the x_1-axis. Altogether, the orientation-preserving rotation under discussion is given by the mapping $v \mapsto A \cdot v$ ($v \in \mathbb{R}^3$) with $A := D_1^{-1} \cdot D_2^{-1} \cdot D_\varphi \cdot D_2 \cdot D_1$. Multiplying out yields

$$A = \begin{pmatrix} \nu_1^2\mu + \cos(\varphi) & \nu_1\nu_2\mu - \nu_3\sin(\varphi) & \nu_1\nu_3\mu + \nu_2\sin(\varphi) \\ \nu_1\nu_2\mu + \nu_3\sin(\varphi) & \nu_2^2\mu + \cos(\varphi) & \nu_2\nu_3\mu - \nu_1\sin(\varphi) \\ \nu_1\nu_3\mu - \nu_2\sin(\varphi) & \nu_2\nu_3\mu + \nu_1\sin(\varphi) & \nu_3^2\mu + \cos(\varphi) \end{pmatrix},$$

where we have set $\mu := 1 - \cos(\varphi)$; this can now be easily decomposed as

$$A = E + \sin(\varphi) \cdot N + (1 - \cos(\varphi)) \cdot N^2$$

with

$$N := \begin{pmatrix} 0 & -\nu_3 & \nu_2 \\ \nu_3 & 0 & -\nu_1 \\ -\nu_2 & \nu_1 & 0 \end{pmatrix},$$

as asserted.

Selected Literature

The following list of books on (elementary) number theory and algebra can serve to fill in some of the gaps in this book's presentation. Some of these books will take the reader much deeper into various topics. The literature on the concept of number and the representation of numbers is of cultural-historical significance, while the works of a historical nature provide insight into the historical development of algebra and number theory. Finally, we offer the interested reader two books on approaches to the teaching of algebra and number theory.

Selected literature for the appendices is listed at the end of the respective appendix.

Literature on Number Theory

[1] D. Burton: *Elementary number theory.* McGraw-Hill Education, 7th edition, 2010.

[2] W. A. Coppel: *Number theory. An introduction to mathematics.* Springer, Berlin Heidelberg New York, 2nd edition, 2009.

[3] G. H. Hardy, E. M. Wright: *An introduction to the theory of numbers.* Oxford University Press, 6th edition, 2008.

[4] H. Hasse: *Number theory.* Translated from the 3rd German edition. Springer, Berlin Heidelberg New York, 1980.

[5] L.-K. Hua: *Intoduction to number theory.* Translated from the Chinese original by P. Shiu. Springer, Berlin Heidelberg New York, 1982.

[6] K. Ireland, M. Rosen: *A classical introduction to modern number theory.* Springer, Berlin Heidelberg New York, 2nd edition, 1990.

[7] F. Jarvis: *Algebraic number theory.* Springer, Cham Heidelberg New York Dordrecht London, 2014.

[8] G. A. Jones, J. M. Jones: *Elementary number theory.* Springer, London, 1998.

[9] M. B. Nathanson: *Elementary methods in number theory.* Springer, Berlin Heidelberg New York, 2000.

[10] I. Niven, H. S. Zuckerman, H. L. Montgomery: *An introduction to the theory of numbers.* John Wiley & Sons, Hoboken, NJ, 5th edition, 2008.

[11] D. Redmond: *Number theory.* Marcel Dekker, New York, 1996.

[12] K. H. Rosen: *Elementary number theory and its applications.* Pearson, Boston, 6th edition, 2010.

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0

[13] W. Sierpinski: *Elementary theory of numbers.* Elsevier, Amsterdam, PWN, Warsaw, 2nd edition, 1988.
[14] A. Weil: *Basic number theory.* Springer, Berlin Heidelberg New York, 3rd edition, 1995.

Literature on Abstract Algebra

[15] M. Artin: *Algebra.* Pearson, Boston, 2nd edition, 2017.
[16] R. Cooke: *Classical algebra: its nature, origins, and uses.* John Wiley & Sons, Hoboken, NJ, 2008.
[17] D. S. Dummit, R. M. Foote: *Abstract algebra.* John Wiley & Sons, Hoboken, NJ, 3rd edition, 2003.
[18] B. Fine, A. M. Gaglione, G. Rosenberger: *Introduction to abstract algebra.* Johns Hopkins University Press, Baltimore, MD, 2014.
[19] J. Gallian: *Contemporary abstract algebra.* Brooks Cole, 9th edition, 2016.
[20] R. S. Irving: *Integers, polynomials, and rings.* Springer, Berlin Heidelberg New York, 2004.
[21] S. Lang: *Algebra.* Springer, Berlin Heidelberg New York, 3rd edition, 2002.
[22] F. Lorenz: *Algebra. Volume I. Fields and Galois theory.* Translated from the 1987 German edition by S. Levy. Springer, Berlin Heidelberg New York, 2006.
[23] W. K. Nicholson: *Introduction to abstract algebra.* John Wiley & Sons, Hoboken, NJ, 4th edition, 2012.
[24] J. J. Rotman: *A first course in abstract algebra.* Pearson, Boston, 3rd edition, 2005.
[25] L. H. Rowen: *Algebra: groups, rings and fields.* A K Peters, Wellesley, MA, 1994.
[26] J. Stillwell: *Elements of algebra: geometry, numbers, equations.* Springer, Berlin Heidelberg New York, 1994.
[27] B. L. van der Waerden: *Algebra.* Volume I. Springer, Berlin Heidelberg New York, 9th edition, 1993.

Literature on the Concept of Number

[28] J. H. Conway, R. K. Guy: *The book of numbers.* Springer Copernicus, New York, 1996.
[29] L. Corry: *A brief history of numbers.* Oxford University Press, Oxford, 2015.
[30] H. Ebbinghaus et al.: *Numbers.* Translated from the 2nd German 1988 edition by H. L. S. Orde. Springer, Berlin Heidelberg New York, 1991.
[31] G. Ifrah: *From one to zero: a universal history of numbers.* Translated from the French original by L. Bair. Penguin Books, New York, 1987.

[32] K. Menninger: *Number words and number symbols: a cultural history of numbers.* Translated from the German revised edition by P. Broneer. Dover, New York, 1992.
[33] R. Taschner: *Numbers at work: a cultural perspective.* Translated from the 2005 German original by O. Binder and D. Sinclair-Jones. A K Peters, Wellesley, MA, 2007.

Literature on the History of Algebra and Number Theory

[34] I. G. Bashmakova, G. S. Smirnova: *The beginnings and evolution of algebra.* Translated from the Russian original by A. Shenitzer. Mathematical Association of America, Washington, DC, 2000.
[35] V. J. Katz, K. H. Parshall: *Taming the unknown: a history of algebra from antiquity to early twentieth century.* Princeton University Press, Princeton, NJ, 2014.
[36] A. Weil: *Number theory.* Birkhäuser Boston, Boston, MA, 1984.

Literature on the Teaching of Algebra and Number Theory

[37] A. Arcavi, P. Drijvers, K. Stacey: *The learning and teaching of algebra.* IMPACT Series, Routledge, 2016.
[38] J. D. Sally, P. J. Sally: *Integers, fractions, and arithmetic: a guide for teachers.* American Mathematical Society, Providence, RI, 2012.

Index

© Springer International Publishing AG 2017
J. Kramer and A.-M. von Pippich, *From Natural Numbers to Quaternions*,
Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-319-69429-0