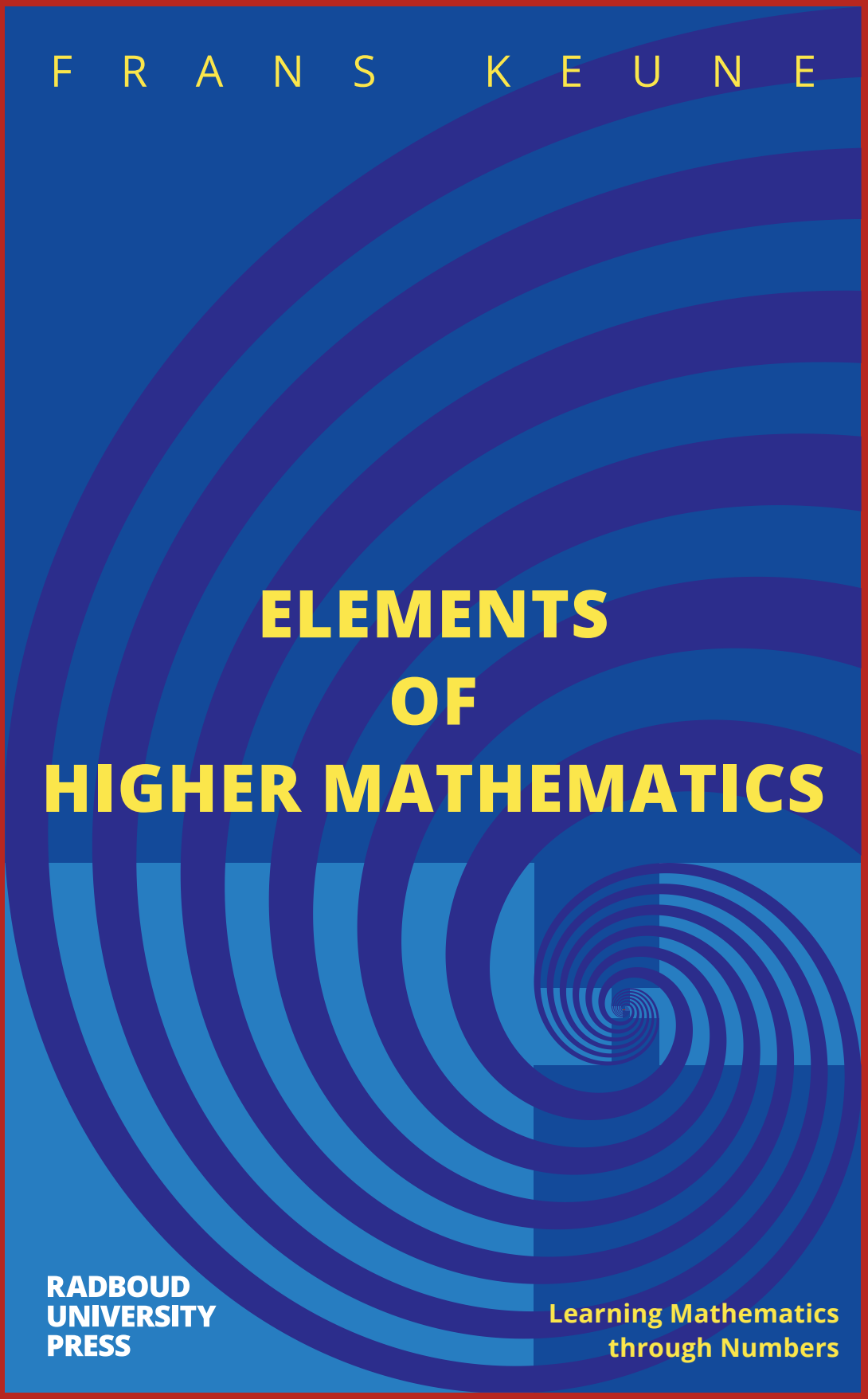
FRANS KEUNE

ELEMENTS OF HIGHER MATHEMATICS



RADBOUD UNIVERSITY PRESS

Learning Mathematics through Numbers

Elements of Higher Mathematics

Learning Mathematics through Numbers

Frans Keune

Originally published as *Getallen, van natuurlijk naar imaginair*,
by Frans Keune. Epsilon Uitgaven, Utrecht 2009

Elements of Higher Mathematics, Learning Mathematics through Numbers

Published by RADBOUD UNIVERSITY PRESS
Postbus 9100, 6500 HA Nijmegen, The Netherlands
www.radbouduniversitypress.nl | radbouduniversitypress@ru.nl

Translation and editing: Frans Keune
Lay out: Frans Keune
Cover design: Frans Keune and Textcetera, The Hague
Cover image: Golden spirals
Print and distribution: Pumbo.nl

Mathematics Subject Classification (2020): 03-01, 05-01, 11-01

ISBN: 9789493296824
DOI: 10.54195/PEAQ9203
Free download at: www.radbouduniversitypress.nl

RADBOUD
UNIVERSITY
PRESS

Preface

This book is primarily intended as a textbook for beginning mathematics students. It does not presume any knowledge of mathematics: it starts with the basics of counting. However, it is certainly helpful having digested some school mathematics: familiarity with some mathematical notions and less fear for formulas.

Large parts of the book originate from lecture notes of a first year course at the Radboud University. It is a mildly edited version of the original Dutch edition "Getallen", which was published by Epsilon Uitgaven in Utrecht, The Netherlands.

The best way to learn mathematics is by doing mathematics. For beginning students it sometimes is a problem determining what to assume when looking for a proof. For the exercises in this textbook this situation does not occur: except for Part I all is built on Peano's axioms for the natural numbers, using the language of intuitive set theory only. Part I describes the way mathematics works: the use of set theory and the relation between language and mathematical entities.

The common thread in the book is the construction of the number system all the way from the natural numbers, via the rationals and the reals to the complex numbers. For the student the advantages of this approach are:

- One learns concepts which are fundamental for all of mathematics.
- The common thread offers a natural way for the introduction of these concepts. It helps to stay motivated during the course.
- One learns to think like a mathematician.
- One obtains insight in the way mathematics is built from simple ideas.
- It helps to decide whether one is fitted for a mathematics study.

For the interested reader extra topics—not covered by the lectures—are included, such as in particular the other possible completions of the rationals, the p-adic numbers. The book contains more than just the construction of the number system: there is also attention for its use, especially in combinatorics, number theory and cryptography, leaving mathematical analysis to the many textbooks for analysis and calculus courses.

I have benefited a lot of comments of students as well as of one of the reviewers. Cian Jameson made suggestions for improving the English in the book. I am grateful to all of them.

Nijmegen, August 2024 *Frans Keune*

Contents

Introduction

For doing mathematics three aspects are important:

1. understanding how mathematics works,
2. understanding some mathematics,
3. creating mathematics.

Having seen large parts of mathematics, for instance for use in other disciplines, does not guarantee that one really understands the way mathematics works. For many people mathematics is a tool and the more the tool is available in software, the less they have to understand. Real understanding assumes an understanding of the way mathematics is organized. Mathematicians try to add something new to mathematical knowledge, a new result or a new proof for an old result. Of course, understanding a lot of mathematics is helpful for doing mathematics. But creativity is on all levels: for instance solving an exercise problem is a deed of creativity.

Higher Mathematics

This book is on 'higher' mathematics. It is the mathematics taught at universities. Higher mathematics is just mathematics. The meaning is not: difficult mathematics. Often, if one says that something is higher mathematics, it means that it is incomprehensible for an ordinary person. The truth of a new result in mathematics is established by logic alone. No mathematics without proofs. On the other hand one has 'school mathematics'. In school mathematics there is the tendency to omit proofs and to be sloppy with definitions of concepts. If the title of the book was 'Elements of Mathematics' the title would not say much and might even give a wrong impression of its contents.

There are five parts. The idea behind each of them is explained next in this introduction. More explanation is on the introductory page in each of the parts. Some concepts are illustrated using the programming language Python. The motivation behind this is explained below.

The book contains too much mathematics for a beginning course in mathematics. Several choices are possible, see below under Course: Introduction to Mathematics.

Overview

Part I: The Rules of the Game. In this part it is explained how mathematics is organized, how it works. It is done by analyzing the puzzle *The Tower of Hanoi*, an invention of the French mathematician Edouard Lucas. Nowadays mathematics is founded on the notion of *set*, introduced by the German mathematician Georg Cantor. In mathematics it is customary to work with sets in an intuitive way. It is closely connected with logic. In this first part of the book knowledge of simple arithmetic is assumed. The assumptions are even further reduced in the next part.

Part II: Foundations. The book series *Elements* of Euclid of Alexandria dates from around 300 BC. For long it was considered to be the basis for all of mathematics. The fundamental notions were geometrical: point, line, circle. Numbers were used, but only in an intuitive way. In this book we start with natural numbers the way Euclid started by giving *axioms* for geometrical objects. The axioms for the natural numbers are *Peano's axioms*, named after the Italian mathematician Giuseppe Peano. The principles of counting are explained and in this part the number system is extended, first with negative numbers and later with fractions.

Part III: Investigations and Applications. Part II is mainly about the foundation of the number system. Now we can investigate the system that has been created by the human mind. For instance the notion of prime number emerges and one may ask simple questions about these numbers and the factorization of integers. Numbers are used for counting. This leads to what is known as combinatorics. Modular arithmetic and in particular the theory of quadratic residues is applied to testing the primality of numbers.

Part IV: Completions. The rational numbers do not suffice, especially not in geometry. In this part the step from rational to real is made. This is the biggest step in the construction of the number system. It is done by completing the number system using the absolute value on the rationals. The method is *analytic*, the previous steps being of a more *algebraic* character. The real numbers are basic for what is known as *calculus*. There are other absolute values on the rationals, for each prime there is one. Completion with respect to these other absolute values leads to totally different number systems.

Part V: Extensions. The step from rational to real is a big one. A more modest step is done by *adjunction* of the root of an algebraic equation, the minimal way of extending a number system such that this root is in the extension. The simplest nontrivial extension is by adjoining the square root of a number which is not a square. This part starts with the construction of the complex numbers, the number system obtained by adjoining the square root of -1.

The complex number system is the end of a chain of constructions. Applications of adjoining square roots in number theory are given. This is the most advanced part of the book.

Course: Introduction to Mathematics

The course I used to give treats the construction of the number system as well as properties of this system. It gives the opportunity to learn the fundamentals of mathematics. This does not include the chapters 14, 15, 18, 20 and 21 containing topics in number theory: quadratic residues, prime tests, p-adic numbers, Hilbert symbols, infinite continued fractions. Some sections of the other chapters can be omitted as well, depending on time and on ones personal preferences. Such a course is ideally given in the first semester of the first year. The non-treated topics are meant as furher reading for those interested.

Python

The computer programming language Python is used to strengthen the formal aspects of mathematics: mathematical notions are clear, they can be handled by computer. An advantage of Python is its clear syntax. Not much is needed for understanding many of the simple Python functions. They are simply straightforward translations of mathematical definitions and procedures. Not every mathematician likes computer programming. The Python sections in the book are not necessary for the mathematics, so one may skip them without any problem. On the other hand, just having a look at the returns of some of the Python functions is already instructive. The Python code can be downloaded from the website of the Radboud University Press. The book 'Think Python' by Alan Downey offers a nice introduction to Python. It is published by O'Reilly and freely downloadable from the website of Green Tea Press.

Further Reading

For those who want more help when making the step from school mathematics to higher mathematics the following books may be useful.

[1] Lara Alcock, *How to Study for a Mathematics Degree*, Oxford University Press, Oxford (UK), 2009.

[2] R. Earl, *Towards Higher Mathematics*: *A Companion*, Cambridge University Press, Cambridge, etc., 2017.

[3] P. Eccles, *An Introduction to Mathematical Reasoning*: *Numbers, Sets and Functions*, Cambridge University Press, Cambridge, etc., 1997.

[4] K. Houston, *How to Think Like a Mathemaician*: *A Companion to Undergraduate Mathematics*, Cambridge University Press, Cambridge, etc., 2009.

The common thread of this textbook is an introducion to mathematics through the construction of the number system. The extras in the book are mainly about the study and the use of numbers: in number theory and in combinatorics. For supplementary reading the following books are advised.

[5] J.H. Conway and R.K. Guy, *The Book of Numbers*, Copernicus, New York, 1996.

[6] R. Courant and H. Robbins, *What is Mathematics?*: *An Elementary Approach to Ideas and Methods*, 2nd ed. (I. Stewart, ed.), Oxford University Press, Oxford UK, 1996.

[7] R Crandall and C. Pomerance, *Prime Numbers*: *A Computational Perspective*, 2nd ed., Springer Verlag, Berlin, etc., 2005.

[8] G.H. Hardy, *A Course of Pure Mathematics*, 10th ed., Cambridge University Press, Cambridge, UK, 1993.

[9] J.-P. Serre, *Cours d'Arithmétique*, Presses Universitaires de France, 1970; English transl. in *A course in arithmetic*, Graduate Texts in Mathematics, vol. 7, Springer-Verlag, New York, 1973.

[10] H. Stark, *An Introduction to Number Theory*, The MIT Press, Cambridge, Mass., 1973.

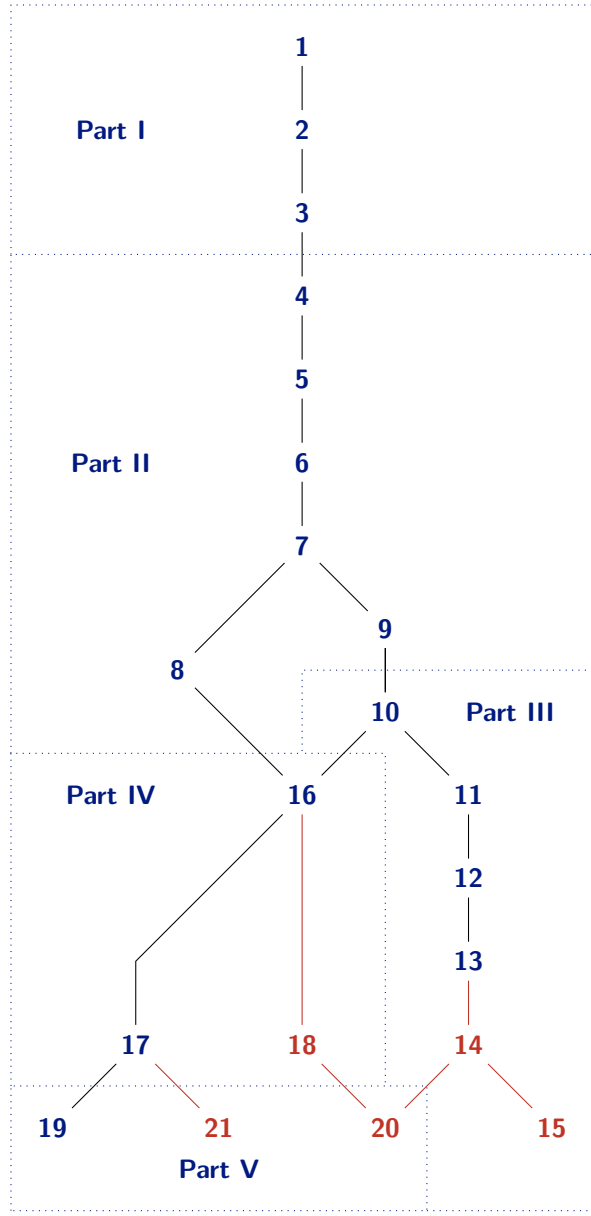
An excellent general introduction to 'higher' mathematics is [6] by Richard Courant and Herbert Robbins. The first edition dates from 1941. Later, in the 90's it was edited and brought up to date by Ian Stewart.

The book [8] by the well-known number theorist G.H. Hardy is a textbook for beginning mathematics students and is mainly focused on mathematical analysis, an important part of mathematics. The first edition dates from 1908 and was edited by the author till 1937.

John Conway and Richard Guy wrote the beautiful book [5] on numbers. Harold Stark, another famous number theorist, wrote a well written easily accessible introduction to number theory [10].

Parts of chapter 20 are based on Jean-Pierre Serre's [9], which however, is written on a much higher level. The well-written book [7] by Robert Crandall and Carl Pomerance describes many algoritms for prime testing and factorization of integers, the subject of chapter 15.

Logical dependence of chapters



The dependence does not apply to the examples given in the chapters.
The red numbers indicate the chapters containing extra topics.

Part I

The Rules of the Game

The mathematics in this book is built from scratch. This is done from Part II onwards. In this introductory part it is explained the way we do this. Characteristic for mathematics is that new results require a proof: they have to be a logical consequence of earlier results. It is inevitable that something has to be accepted in advance. In this book this will be some very simple properties of natural numbers, the numbers 0, 1, 2, ...

The logic will be not more than what is used in every day life. Only, the language is made more precise: for example we have to agree on the meaning of the word 'or'. Also it is important that we know what we are talking about, what the mathematical objects are. For this we use the notion of *set*. We will deal with sets in an intuitive way. This is made more precise in chapter 2.

The notion of *graph* in chapter 3 serves as a first example of an abstract mathematical structure. Abstraction is characteristic for mathematics. Irrelevant details are omitted and abstract structures remain. We try to clarify this with a puzzle: the Tower of Hanoi, described in chapter 1. Of course, there are more important things than puzzles. Here it's the principle that matters.

If the language is very precise, it is suited for communication with a computer. A good understanding of an abstract structure enables us to instruct the computer doing tasks which for humans are dull and time consuming.

1 The Tower of Hanoi

The *Tower of Hanoi* is a puzzle. It was invented in 1883 by the French mathematician Edouard Lucas. The puzzle consists of three vertical pegs and a number of discs. A hole in the middle of each disc makes it possible to place them on the pegs. The discs have varying diameters and are placed in order of decreasing size on one of the pegs, see Figure 1.1. The aim is to transfer all discs to one of the two other pegs by making moves. A move is the transfer of one of the top discs from one peg to one of the two others. It is not allowed to place a disc on top of a smaller one. It is not so easy to provide a full proof description of the puzzle in plain English. Usually an appeal is made on the reader's intuition and also a picture like the one in Figure 1.1 makes it more understandable. In daily life that suffices.

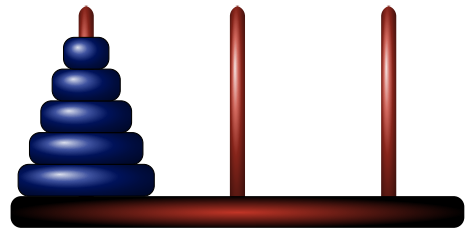


Figure 1.1: Starting position (5 discs)

There are three pegs. We label them with the digits '1', '2' and '3'. A position of the puzzle will be denoted by a *word* (or *string*) in these three digits: the first digit indicates on which peg the largest disc is situated, the next digit indicates where the second largest is, etc. The position in Figure 1.2 is denoted by the word '11232'. It is a word of 5 digits, since it is a puzzle with 5 discs.

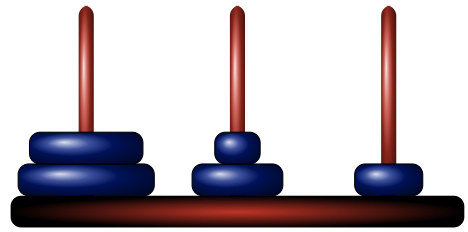


Figure 1.2: The position 11232

Notice how these words of digits are used:

- 1232321 is a position of the Tower of Hanoi in the case of seven discs,
- '1232321' is a word of seven tokens (digits in this case).

Compare this with sentences like:

- John is my brother,
- 'John' consists of four letters.

F. Edouard Lucas (Amiens 1842 – Paris 1891)

Lucas published his puzzle under the name professor N. Claus (of Siam), mandarin of the College of Li-Sou-Stian. He had made up a nice story. The game was found in the writings of the mandarin Fer-Fer-Tam-Tam which by decree of the Chinese government had to be published in the near future. In Japan, China and Tonkin the discs were of porcelain. As we will see, for the solution of a game with 64 discs 18 446 744 073 709 551 615 moves are needed. According to an old legend priests in the temple of Benares are solving the puzzle with 64 golden discs with diamonds. As soon as all the discs are transferred to another peg, the world will vanish.

A move is the transfer of a disc from one peg to another. For the corresponding words this means the replacement of one of the digits by another. The move is only allowed if:

- the disc to be transferred is on top: the same digit has no occurrence in the word to the right of the digit to be replaced,
- the disc is not placed on top of a smaller one: neither has the new digit an occurrence in the word to the right of the digit to be replaced.

  In position 11232 three moves are possible:

  - the smallest disc is transferred from peg 2 to peg 1,
  - the smallest disc is transferred from peg 2 to peg 3,
  - the second smallest disc is transferred from peg 3 to peg 1.

  The new positions after these moves are respectively 11231, 11233 and 11212, see Figure 1.3.

It is irrelevant whether the puzzle is made of wood, porcelain or gold. Many implementations are possible. It can be without pegs, with matches of different sizes instead of discs, or with the rule that discs have to be placed on smaller ones. The words of digits for the positions and the allowed moves are applicable to all these variations. It makes communication about the game possible, irrespective of its implementation. In a way the Tower of Hanoi with a given number of discs is an abstract notion: using our imagination there is no need to have it physically realized. Of course it is a bit silly, but there is no problem considering a Tower of Hanoi with a billion discs. The notation for positions is far more efficient than the representation of positions by figures like in Figure 1.2 and Figure 1.3. Even for not so large number of discs, say 64 discs, the use of pictures would be far more complicated and far less understandable.

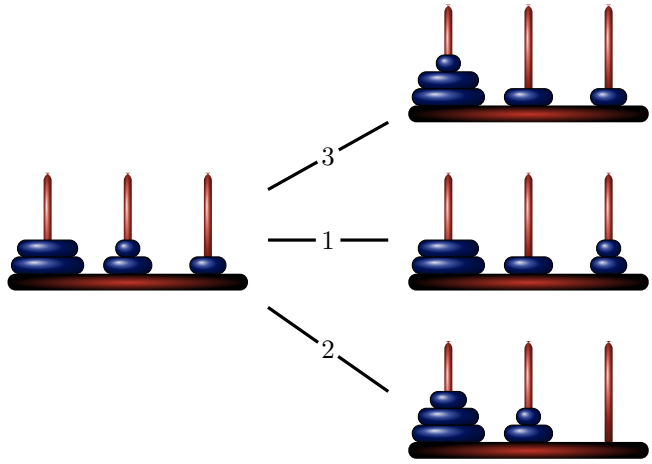


Figure 1.3: Three possible moves

When all discs are on one single peg, two moves are possible. In all other cases the number of possible moves is three. These moves can be described in various ways. Here we will describe a move by giving the initial position together with the peg that is not involved in the move. Such a move does not exist for the position $111\ldots111$ and the peg 1, and similarly for the two other cases in which all discs are on one single peg. In all other cases it describes a unique move. In Figure 1.3 the three moves are indicated this way.

Python

The notation for positions of the Tower of Hanoi enables us to communicate about the puzzle with a computer. Computer languages know the data structure *string*. Strings of the symbols `'1'`, `'2'` and `'3'` can be used to represent positions of the puzzle and moves can be programmed.

In the module `hanoi.py` we keep Python functions related to the Tower of Hanoi. Thus they are available through the command `from hanoi import *`. The function

```
replace(i,word,char)
```

returns the string obtained by replacing the character with index `i` in the string `word` by the character `char`. In Python a character in a string can not be directly substituted by another character: strings are *immutable*. Here the new string is obtained by assembling parts of the old string with the character `char` in between. In Python concatenation of strings is done with `+`. If `i` is not in the range of the

indices of the string `word` the returned string is chosen here to be just the unchanged `word`.

hanoi.py
```
def replace(i, word, char):
    if i not in range(len(word)): return word
    else: return word[:i] + char + word[i+1:]
```

Use of the function `replace`:

```
>>> from hanoi import *
>>> replace(2,'eer7r7eabvey', 'P')
'eeP7r7eabvey'
>>> replace(-1,'eer7r7eabvey', 'P')
'eer7r7eabvey'
>>> replace(17,'eer7r7eabvey', 'P')
'eer7r7eabvey'
```

The function `nextposition(peg,position)` returns the position obtained from the position `position` with the move that does not involve the peg `peg`.

hanoi.py
```
def nextposition(peg, position):
    i1 = position.rfind('1')
    i2 = position.rfind('2')
    i3 = position.rfind('3')
    if peg == '1':
        if i2 < i3: return replace(i3, position, '2')
        else: return replace(i2, position, '3')
    if peg == '2':
        if i1 < i3: return replace(i3, position, '1')
        else: return replace(i1, position, '3')
    if peg == '3':
        if i1 < i2: return replace(i2, position, '1')
        else: return replace(i1, position, '2')
```

For example:

```
>>> nextposition('3', '123331113233')
'123331113133'
>>> nextposition('1', '123331113233')
'123331113232'
>>> nextposition('2', '123331113233')
'123331113231'
>>> nextposition('2','222222222222')
'222222222222'
```

Note that in Python `word.rfind(char)` returns `-1` if the character `char` does not occur in the string `word`.

2 Intuitive Set Theory

In mathematics the only way to establish new results is by logical reasoning. It is reasoning about properties of objects, just the way we do this in every day life. In this chapter the reasoning is about the Tower of Hanoi, the positions of discs and the moves of discs. Here we use simple counting principles. This in contrast to the next chapters, where it is shown how arithmetic is based on some simple assumptions.

2.1 Sets

The Tower of Hanoi with three discs has 27 positions:

$$\begin{array}{ccccccccc} 111, & 112, & 113, & 121, & 122, & 123, & 131, & 132, & 133, \\ 211, & 212, & 213, & 221, & 222, & 223, & 231, & 232, & 233, \\ 311, & 312, & 313, & 321, & 322, & 323, & 331, & 332, & 333. \end{array}$$

We will see all these positions of the Tower of Hanoi with three discs together as one entity, an entity consisting of 27 objects. In mathematics the terms *'set'* and *'element'* are used. The 27 positions of the Tower of Hanoi with three discs are the elements of a set, the set of all these positions. This set is denoted as follows:

$$\{\, 111, 112, 113, 121, 122, 123, 131, 132, 133, 211, 212, 213, 221, 222, \\ 223, 231, 232, 233, 311, 312, 313, 321, 322, 323, 331, 332, 333 \,\}\,.$$

It is denoted as a comma separated enumeration of its elements, enclosed in braces. Let's denote the set of all positions of the Tower of Hanoi with n discs by $V(n)$. Then the above set is the set $V(3)$. So

$$121 \text{ is an element of } V(3)$$

is just another way of saying that

$$121 \text{ is a position of the Tower of Hanoi with three discs.}$$

To indicate that x is an element of a set S, the symbol $\in$ is used: $x \in S$. So for

Georg Cantor (St. Petersburg 1845 – Halle 1918)

The notion of set was originated by Cantor. He introduced new infinite numbers to denote the number of elements of an infinite set (the *cardinal number* of the set, see also page 386).

$$121 \text{ is an element of } V(3)$$

we have the shorter expression

$$121 \in V(3).$$

Note that this still is an ordinary sentence: 121 and $V(3)$ are names of two objects and the $\in$ tells that the first object is an element of the second. It is a *declarative sentence*, i.e. a sentence that makes a statement. To indicate that something is not an element of a set one uses the symbol $\notin$, e.g.

$$1332 \notin V(3),$$

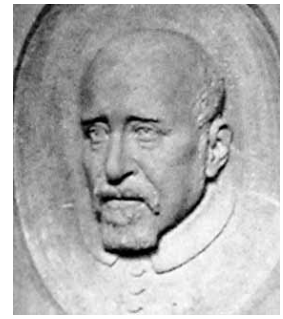
also a declarative statement. In logic declarative statements are usually called *propositions*.

It is useful—as we did above—to consider the totality of all positions of the Tower of Hanoi as one single object. Such an object can have a name (like $V(3)$), which can be used for communication. Objects can have properties: a property of $V(3)$ for example is that it has 27 elements. Sets can also occur as elements of other sets. In the paragraph below and in the next two sections we will see some examples of this phenomenon. The number of elements of a set we usually denote using #: the number of elements of S is $\#(S)$. The number of elements of a finite set is a natural number. In Chapter 4 natural numbers are treated and in Chapter 5 there is more about the notions 'number' and 'finite'.

The Tower of Hanoi puzzle is determined by the totality of all its positions and all pairs of positions that are connected by one single move. For every number n of discs these pairs form a set $E(n)$. Note that the elements of $E(n)$ are sets themselves, in this case sets of two elements. The set $E(2)$ for example is the set

Robert Recorde (Tenby 1510 – London 1558)

Robert Recorde studied medicine in Oxford and Cambridge. He practised medicine in London. He wrote books on medicine, astronomy and mathematics. The last were algebra textbooks written in English: *The Grounde of Artes* and *The Whetstone of Witte*. He introduced the symbol =, but it took two centuries before it was generally accepted.



$$\{\,\{11,12\},\{11,13\},\{12,13\},\{21,22\},\{21,23\},\{22,23\},\\ \{31,32\},\{31,33\},\{32,33\},\{12,32\},\{13,23\},\{21,31\}\,\}\,.$$

The set $E(2)$ has 12 elements.

2.2 Equality

In the notation of a set based on the enumeration of its elements, the order of the elements is irrelevant and multiple occurrences make no difference:

$$\{0,1\},\ \{1,0\},\ \{0,1,0\} \text{ and } \{1,0,1,0,1\}$$

all denote the same set, the set with the numbers 0 an 1 as its members. A set is thought of being completely determined by its elements. Sets having the same elements are equal. In mathematics objects (sets) are equal if there is no distinction between them whatsoever.

2.1 Notation. Equality of mathematical objects A and B is denoted by $A = B$. The symbol = for equality was introduced by the Welsh mathematician **Robert Recorde** for use in equations.

So for instance

$$\{0,1\} = \{1,0\} = \{0,1,0\} = \{1,0,1,0,1\}.$$

These are five different ways to denote the same set.

> Note that in programming languages = often has a different meaning: in Python it is used to assign a value to a variable: `a = 5`, the value 5 is assigned to the variable `a`. For equality == is used:

```
>>> a = 5
>>> a
5
>>> 2 + 3 == 5
True
>>> 5 == 6
False
```

2.3 Properties and Subsets

Some pairs of different positions of the Tower of Hanoi with n elements represent a move and for $n > 1$ there are pairs which do not. Let X be a pair of different positions. Then either X represents a move or it does not. Let's abbreviate the sentence

'X represents a move' to '$P(X)$'.

Thus the P stands for a property a pair of different positions may have: $P(X)$ is true if X represents a move and otherwise $P(X)$ is not true. Let $F(n)$ denote the set of all pairs of different positions. So, in other words, for each X we have a proposition $P(X)$ depending on X. The set $E(n)$ of all different pairs of positions which represent a move can be given as follows:

$$E(n) = \{\, X \in F(n) \mid P(X) \,\}.$$

This is the notation for the set of elements left of the symbol $|$, which have the property P. Using this notation, the set $F(n)$ can be given by

$$F(n) = \{\, \{x, y\} \mid x, y \in V(n) \text{ and } x \neq y \,\}.$$

2.2 Definition. Let A and B be sets. Then B is a *subset* of A if every element of B is also an element of A. Notation: $B \subseteq A$. We say that the set B is *contained* in the set A and also that the set A *contains* the set B.

This is the first definition. In a definition a new notion is introduced. It is described in terms of notions that are already known. Here the new notion is: subset. Notions already known are: set, element. Often a notation is introduced as well; here: $B \subseteq A$. The notation '$B \subseteq A$' is short hand for 'Every element of B is an element of A'. Sometimes there are variations in the terminology, like here 'B is contained in A' and 'A contains B'.

A definition often starts with a sentence describing the context, a sentence telling what it is about. Here that is the sentence: 'Let A and B be sets.' Subsequently the new notion is introduced.

An example of a subset: $E(n) \subseteq F(n)$, every pair representing a move is a pair of different positions. For each $X \in E(n)$ we have another example: $X \subseteq V(n)$.

Every property P that elements of a set A can have determines a subset of A, the set of all $a \in A$ such that $P(a)$:

$$\{\, a \in A \mid P(a) \,\} \subseteq A.$$

Conversely, given a subset B of a set A, the set B is given by a property for elements a of A, namely the property P stating for $a \in A$ that $a \in B$. Properties of elements of a set A are said to be *equivalent* if they determine the same subset of A.

If all elements of a set A are also elements of a set B and, conversely, all elements of B are elements of A, then both sets have the same elements, that is they are equal. So the assertion

$$A \subseteq B \quad \text{and} \quad B \subseteq A$$

comes down to (is *equivalent* to):

$$A = B.$$

We allow a set to have no elements. Such a set is a subset of any other set. If Z_1 and Z_2 are both sets without elements, then $Z_1 \subseteq Z_2$ and $Z_2 \subseteq Z_1$, and so $Z_1 = Z_2$. As a consequence there exists just one set without elements, so we may speak of *the* set without elements.

2.3 Definition. The set without elements is called the *empty set*. It is denoted by $\emptyset$.

A set of three elements, say the set $\{1, 2, 3\}$, has eight subsets: $\{1, 2, 3\}$, $\{2, 3\}$, $\{1, 3\}$, $\{1, 2\}$, $\{1\}$, $\{2\}$, $\{3\}$ and $\emptyset$. Note that both the set itself and the empty set are subsets.

2.4 New Sets from Old

When talking about a property objects may have, it is convenient to look at the totality of all objects with the given property. It helps reasoning with properties. Combining properties leads to constructions with (sub)sets.

Properties P and Q for objects x of, say a given set U, leads to the new property which states that the properties both hold for x, that is $P(x)$ and $Q(x)$.

2.4 Definition. Let A and B be sets. By $A \cap B$ we denote the *intersection* of A and B, the set of the elements that A and B have in common:

$$A \cap B = \{\, x \mid x \in A \text{ and } x \in B \,\}.$$

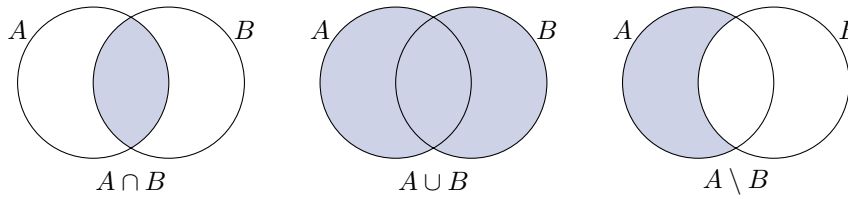


Figure 2.1: Intersection, union and difference of A and B

Note that in all cases there is an intersection, even if the sets have no elements in common: $\{1,2\} \cap \{3,4\} = \emptyset$. That is one of the advantages of having a thing like the empty set. Is the intersection empty, then we say that the sets are *disjoint*.

Just as P and Q leads to '$P(x)$ and $Q(x)$', it also leads to '$P(x)$ or $Q(x)$'. In mathematics 'or' always is the inclusive or: it is not excluded that both assertions hold.

2.5 Definition. Let A and B be sets. By $A \cup B$ we denote the *union* of A and B, the set of the elements that are in at least one of both sets:

$$\{ x \mid x \in A \text{ or } x \in B \}.$$

Clearly for sets A, B and C we have

$$(A \cap B) \cap C = A \cap (B \cap C) \quad \text{and} \quad (A \cup B) \cup C = A \cup (B \cup C).$$

The operations 'taking the intersection' and 'taking the union' are said to be *associative*. In case of an associative operation in expressions like those above omitting parentheses causes no ambiguity and is therefore common practice.

2.6 Definition. Let A and B be sets. The *difference* $A \setminus B$ is the set of the elements of A which are not in B:

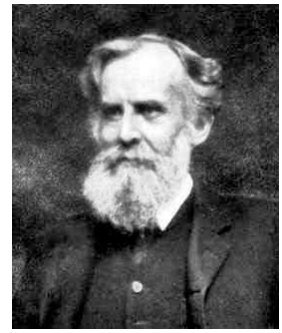
$$A \setminus B = \{ x \mid x \in A \text{ and } x \notin B \}.$$

If $B \subseteq A$, then $A \setminus B$ is also called the *complement* of B in A. For A fixed the complement might also be denoted by B^c or as B'.

Visualizing the sets A and B as discs in the plane, the sets $A \cap B$, $A \cup B$ and $A \setminus B$ can be indicated as in Figure 2.1. This visualization is helpful when reasoning with unions, intersections and differences of sets.. They are known as *Venn diagrams*.

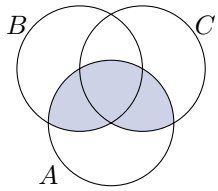
John Venn (Hull 1842 – Cambridge 1923)

John Venn was an English mathematician best known for the Venn diagrams. Though they are named after him, he was not the first who made such diagrams. However, he was the first to formalize their usage.

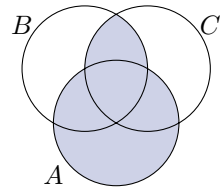


2.7 Rules and logic. Using $\cap$, $\cup$ and $\setminus$ new sets can be made from old. Starting with sets A, B and C various sets can be made. The following rules hold, see Figure 2.2:

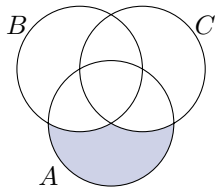
a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,
b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
c) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$,
d) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.



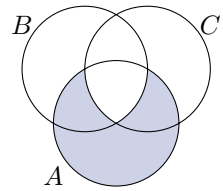
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$



$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$



$A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$



$A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$

Figure 2.2: Rules for intersection, union and difference

For a given object x put p, q and r for respectively the propositions $x \in A$, $x \in B$ and $x \in C$. Then the rules express the following for the object x:

a) p and (q or r)	is equivalent to	(p and q) or (p and r),
b) p or (q and r)	is equivalent to	(p or q) and (p or r),
c) p and not (q or r)	is equivalent to	(p and not q) and (p and not r),
d) p and not (q and r)	is equivalent to	(p and not q) or (p and not r).

Truth tables

A way of understanding the equivalence of propositions is by verifying all possible cases. For each of the propositions p, q and r there are two possibilities: either it is true or it is not. All together for the expressions above there are eight cases.

Given propositions p and q new propositions can be formed using 'and', 'or', 'not', 'if, then' and 'if and only if' (= 'is equivalent to' and often abbreviated as 'iff'). The truth of these new propositions is determined by the truth of p and q. If we indicate by a 1 that a proposition is true and by a 0 that it is not, then we can form a *truth table*, a table that shows how the truth of the new proposition is determined by the truth of p and q.

p	not p
0	1
1	0

p	q	p and q	p or q	if p, then q	p iff q
0	0	0	0	1	1
0	1	0	1	1	0
1	0	0	1	0	0
1	1	1	1	1	1

Logical symbols may be used:

not p:	$\neg p$
p and q:	$p \wedge q$
p or q:	$p \vee q$
if p, then q:	$p \Rightarrow q$
p if and only if q:	$p \Leftrightarrow q$

In mathematical texts usually only the last two logical symbols are used, and mostly not very frequently.

As an example of the use of truth tables we verify the logical equivalence of the propositions $p \wedge (q \vee r)$ and $(p \wedge q) \vee (p \wedge r)$:

p	q	r	$q \vee r$	$p \wedge (q \vee r)$	$p \wedge q$	$p \wedge r$	$(p \wedge q) \vee (p \wedge r)$
0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0
0	1	0	1	0	0	0	0
0	1	1	1	0	0	0	0
1	0	0	0	0	0	0	0
1	0	1	1	1	0	1	1
1	1	0	1	1	1	0	1
1	1	1	1	1	1	1	1

The meaning of the logical connectives is given by their truth table. For understanding the truth of 'if p, then q' for untrue p and true q some explanation may be helpful. Two reasons:

- Let A and B be subsets of a set U such that $A \subseteq B$. The proposition $A \subseteq B$ now has the description

  $$\text{for all } x \in U\text{: if } x \in A\text{, then } x \in B.$$

  In particular, also in case $x \in U \setminus A$, the proposition 'if $x \in A$, then $x \in B$' is considered to be true.
- If there exists some reasoning which leads from a proposition p to a proposition q, you want 'if p, then q' to be true. Take for p the proposition $0 = 1$, then also $1 = 0$. So $0+1 = 1+0$. From a false p we derived a true proposition.

Power sets and a paradox

Sets can be formed that have sets as their elements. To get used to this idea we consider here the set of *all* subsets of a set:

2.8 Definition. Let A be a set. The set whose elements are the subsets of A is called the *power set* of A. This set is denoted by $\mathcal{P}(A)$. Thus:

$$\mathcal{P}(A) = \{\, U \mid U \subseteq A \,\}.$$

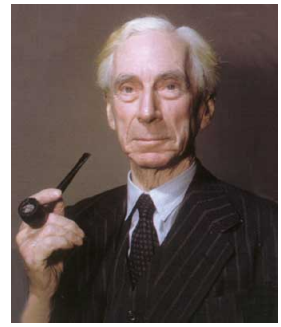
We have

$$\begin{aligned}
\mathcal{P}(\{1,2,3\}) &= \{\{1,2,3\},\{2,3\},\{1,3\},\{1,2\},\{1\},\{2\},\{3\},\emptyset\} \\
\mathcal{P}(\{1,2\}) &= \{\{1,2\},\{1\},\{2\},\emptyset\} \\
\mathcal{P}(\{1\}) &= \{\{1\},\emptyset\} \\
\mathcal{P}(\emptyset) &= \{\emptyset\}.
\end{aligned}$$

If $\#(A) = n$, then A has 2^n subsets, or equivalently $\#(\mathcal{P}(A)) = 2^n$: a subset U of A is given by specifying for each of the n elements of A whether it is an element of U or not. See also section 5.9.

Lord Bertrand Russell (Ravenscroft 1872 – Penrhyndeudraeth 1970)

Russell contributed to the foundations of mathematics and was known as a philosopher. Together with A.N. Whitehead he wrote the book *Principia Mathematica* on the logical foundations of mathematics. He was an activist, originally as a pacifist. The Russell tribunal for war crimes in Vietnam was named after him.



The theory of sets as introduced by **Cantor** contained contradictions. Restrictions to the formation of sets were needed to avoid contradictions. *Russell's paradox* is a clear example of a contradiction:

> Let X be the set of all sets. This is a very very large set. For X it holds that $X \in X$, because X itself is a set as well. That already is strange. Let's take the set of all sets which do not have this strange property:
>
> $$Z = \{ Y \mid Y \notin Y \}.$$
>
> What is the case: $Z \in Z$ or $Z \notin Z$? If $Z \in Z$, then $Z \notin Z$. And if $Z \notin Z$, then $Z \notin Z$ does not hold, so $Z \in Z$.

So there can't be such a thing as the set of all sets. Russell's paradox teaches us to be careful with the formation of sets. In this book we base ourselves on *intuitive set theory*. That is what mathematicians usually do. Because set theory plays the role of foundation for the whole of mathematics it is important that paradoxes are avoided. At the same time the theory should be flexible enough to build mathematics as we like it to be. This is the purpose of *axiomatic* set theory. For our purposes intuitive set theory suffices: we refrain from wild formations of sets.

Exercises

1. Determine $\#(V(n))$, i.e. the number of positions of the Tower of Hanoi with n discs.
2. As the previous exercise, but now for $\#(E(n))$.
3. Let A and B be sets. Show that $A \cap B = B$ is equivalent to $B \subseteq A$. Also show that $A \cup B = A$ is equivalent to $B \subseteq A$.

4. Let P and Q be propositions. Show that the following propositions are equivalent:

$$P \Rightarrow Q, \quad (\neg P) \vee Q, \quad \neg(P \wedge (\neg Q)).$$

5. We define the *symmetric difference* $U \div V$ of sets U and V as follows

$$U \div V = (U \setminus V) \cup (V \setminus U).$$

Let A be a set and U, V and W subsets of A.

(i) Show that $U \div V = (U \cup V) \setminus (U \cap V)$.

(ii) Show that $(U \div V) \div W = U \div (V \div W)$.

(iii) Show that $U \div \emptyset = U$, $U \div U = \emptyset$.

(iv) Show that there is a unique $X \in \mathcal{P}(A)$ such that $U \div X = V$.

6. Let $V = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

(i) How many subsets does V have?

Let W be the set of all words of length 10 formed with the digits 0 and 1, for example 0101010101 and 0000000000.

(ii) Determine $\#(W)$.

(iii) What is the connection between $\mathcal{P}(V)$ and W ?

7. Let A and B sets. Show that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$ is equivalent to $A \subseteq B$. If $\mathcal{P}(A) = \mathcal{P}(B)$, does $A = B$ hold as well?

8. Let A and B be sets.

(i) Show that $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cap \mathcal{P}(B)$.

(ii) If A and B are disjoint, are $\mathcal{P}(A)$ and $\mathcal{P}(B)$ then disjoint as well?

(iii) Give sets A and B such that $\mathcal{P}(A \cup B) \neq \mathcal{P}(A) \cup \mathcal{P}(B)$.

9. Let V be a set of n elements and W the set of all subsets D of V with $\#(D) = 2$. Determine $\#(W)$.

Directions for doing the exercises

a) If in an exercise it is asked to show something, then the answer consists of a reasoning. Formulate the solution in ordinary English.

b) While reasoning, use common sense. Act as if abstract entities like numbers and sets are ordinary objects like apples and oranges.

c) When it is asked to show something for every set A, assume some arbitrary set A being given. Then use that A is a set and that this is all we know about A.

d) In the exercises 1, 2, and 9 it is asked to determine a number depending on an arbitrary number n. The answer is a formula containing n. Clarify as well as possible the correctness of that formula. Anything helpful, like drawing pictures, is allowed.

3 Structure

It is the abstractness of mathematics that allows for mathematical models of real phenomena. These can be technical, economical or from natural science, but also from daily life. Since a mathematical model is an abstract structure, its properties can not be established by observation. Logical reasoning is the only way. Subsequently, properties of the model can be interpreted in terms of the original real phenomena. This is the general situation when mathematics is applied. In this chapter we consider just one type of such an abstract mathematical structure: a graph.

3.1 Graphs

3.1 Example. Given a company of eight persons, let's assume that for each pair of them there are two mutually exclusive possibilities: they know each other or they don't. In order to indicate what the situation is, we can enumerate all pairs of persons that know each other. Let's assume that we have labelled the persons with the numbers 1 to 8. The enumeration can be something like $\{1,2\}$, $\{1,4\}$, $\{1,5\}$, $\{2,3\}$, $\{2,6\}$, $\{3,4\}$, $\{3,7\}$, $\{4,8\}$, $\{5,6\}$, $\{5,8\}$, $\{6,7\}$, $\{7,8\}$. That is for instance all we need when we want an answer to the question: is there for every pair of persons a person they both know? The situation can be further clarified by drawing a picture: dots for the persons (or numbers) and lines connecting persons that know each other, see Figure 3.1. It is clear that for the persons 1 and 7 there is no person in the company they both know.

The picture provides a clear overview of the whole structure. Taking the vertices of a cube and as a property of a pair of vertices 'being connected by an edge of the cube', will result in the same picture. And the same holds for the sides of an octahedron and the property 'having an edge of the octahedron in common'. This structure is an example of a graph. A graph consists of a set (of vertices of the graph) together with a set of subsets of two elements each (the edges of the graph). This is typical for a mathematical structure: a set with

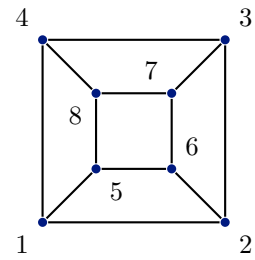


Figure 3.1: Picture of a graph

something extra satisfying certain conditions. We will give a more precise definition.

3.2 Definition. A *graph* G $(= (V, E))$ is a finite set V together with a set E of subsets of V each having exactly two elements. The elements of V are called the *vertices* of the graph, those of E the *edges*.

Thus in the graph described in Example 3.1 we have $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and $E = \{\{1, 2\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 6\}, \{3, 4\}, \{3, 7\}, \{4, 8\}, \{5, 6\}, \{5, 8\}, \{6, 7\}, \{7, 8\}\}$.

The notion of 'graph' is completely described in terms of sets. A graph G consists of two sets: the set of vertices of G and the set of edges of G. Edges are just pairs of (different) vertices. And a vertex is nothing else then an element of a set, the set of vertices. That is all. What the vertices actually are is irrelevant. This is a first example of an abstract mathematical structure. Later we will see other examples: partitions, groups, rings, fields, ordered sets, ...

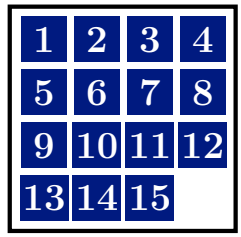


Figure 3.2: Sliding puzzle

3.3 Example (14-15 puzzle). A well-known puzzle consists of fifteen square blocks numbered 1 up to 15 on a 4 by 4 board of square fields. The aim is to slide the blocks from an arbitrary position on 15 of the 16 squares in such a way that they are in the right order with the field below right being empty. One might consider the graph of this puzzle: the vertices are all the positions of the puzzle and edges indicating single moves between two of the positions. We are not going to make a picture of this graph: it has $16 \cdot 15 \cdot 14 \cdots 3 \cdot 2$ vertices and even more edges. This puzzle, with only the blocks 14 and 15 interchanged, was introduced as the '14-15 puzzle' by **Sam Loyd** in 1878. For analyzing the puzzle it is helpful to discover more structure than just the structure of a graph. We will do so in chapter 12, section 12.5.

Figure 3.3: Mini sliding puzzle

An extremely simplified version of the puzzle consists of 3 blocks on a 2 by 2 board. It has only 24 positions. From each position exactly two others can be reached in one move. While sliding either you slide backwards or you slide to a unique next position. In any situation it is obvious whether it is possible to move the blocks to a given position. A position can be denoted by placing the numbers 0, 1, 2 and 3 in some order, for example 2103 denotes the position given in Figure 3.3: 2 in field 1, 1 in field 2 and 3 in field 4. The 0 is used to indicate which field is empty. In this notation a move amounts to interchanging the 0 with one of the other numbers, but this not allowed when the two numbers are both in the middle or both at the end. In Figure 3.4 a picture of this graph is shown. As you see the positions come in two kinds: in half of the positions the right order can be reached. The same holds for the sliding puzzle with 15 blocks, but that result is much harder to obtain.

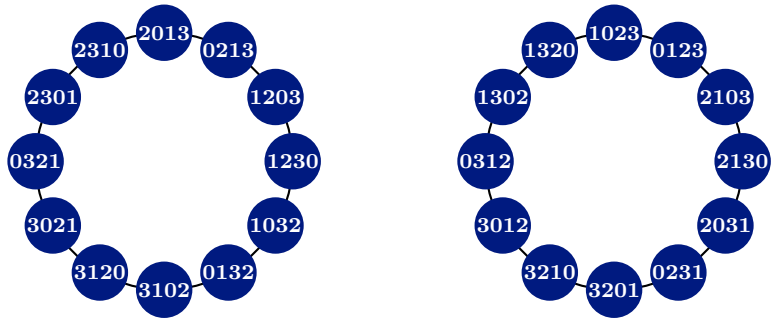


Figure 3.4: Graph of the mini sliding puzzle

3.4 Example (The Tower of Hanoi). To the Tower of Hanoi with n discs there is an associated graph $H(n)$: a vertex set $V(n)$ together with a set $E(n)$ of edges. (Thus $H(n) = (V(n), E(n))$.) Figure 3.5 is a picture of the graph $H(3)$. In the next section the way it is constructed will be explained. This graph is *connected*, that is one can go from any vertex to any other vertex by moving along the edges. If the vertex below left is 111 and if the top vertex is 333, then one sees in the picture of the graph that it is possible to go from position 111 to position 333 in 7 moves. Since the graph is connected any position of the discs can be reached from any other position.

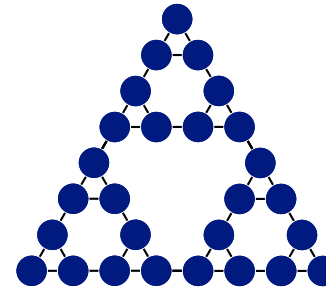


Figure 3.5: The graph $H(3)$

3.2 The Solution of the Tower of Hanoi

The graph $H(1)$ associated to the Tower of Hanoi with just one disc is quite trivial, see Figure 3.6, where the positions are given by pictures, respectively by codes.

In the Tower of Hanoi with two discs there are six moves with the largest disc: this disc is on one of the three pegs and it can move to the peg with no disc. If the largest disc remains fixed we are in fact dealing with three Towers of Hanoi with just one disc, see the top of Figure 3.7. In the bottom left codes for the positions are used and all possible moves are indicated: it is a picture of the graph $H(2)$. The graph in the bottom right is another picture of $H(2)$ using mirror images of the small triangles. In general: leaving the largest disc fixed in the Tower of Hanoi with $n + 1$ discs means you are dealing with three Towers of Hanoi with n discs.

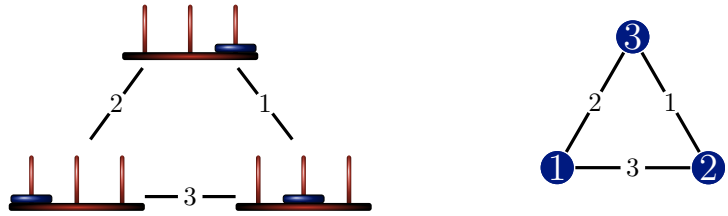


Figure 3.6: Moves of the Tower of Hanoi with one disc

Adding the six moves of the largest disc completes the Tower of Hanoi with $n+1$ discs.

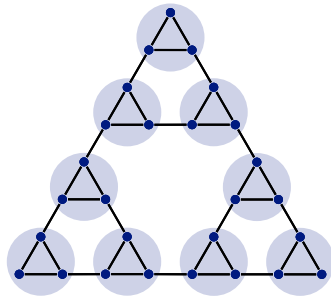


Figure 3.8: 9 subgraphs of $H(3)$

The edges of $H(3)$ corresponding to moves of the smallest disc constitute 9 subgraphs of 3 vertices each, see Figure 3.8. Moving in the graph $H(3)$ from the vertex below left (position 111) to the top vertex (position 333), consists of 7 single moves. Alternately the smallest disc and the unique other possible disc are moved. The moves of the smallest disc are from peg 1 to peg 3, from peg 3 to peg 2, or from peg 2 to peg 1. That determines which of the two possible moves of the smallest disc have to be made. In case of 4 discs the moves of the smallest disc are in the opposite direction, if the goal is position 3333, otherwise you will end in position 2222. That is related to the graph $H(4)$ being obtained out of three copies of $H(3)$ using reflections. It is clear that the direction of the moves of the smallest disc is determined by the parity of the number of discs. This solution is one of the easiest to remember. The solution of the Tower of Hanoi with 2 discs is easily seen in Figure 3.7, which is a picture of the graph $H(2)$. From 11 one comes to 22 with three moves: first move 2, then 3 and finally 1. This succession can be noted by 231. From 22 one arrives at 33 by 312.

For the solution of the Tower of Hanoi with 3 discs first go from 111 to 122, next by move 2 we arrive at 322 and finally go to 333. The resulting succession of moves is 2312312. The first three moves are those of $H(2)$ from 11 to 22, next move 2 and finally three moves of $H(2)$ from 22 to 33.

For $H(4)$ this way one obtains 321321321321321. For $H(n)$ with n even one goes from $111\ldots111$ to $333\ldots333$ with $321321321\ldots$ and for odd n this is done by $231231231231\ldots$.

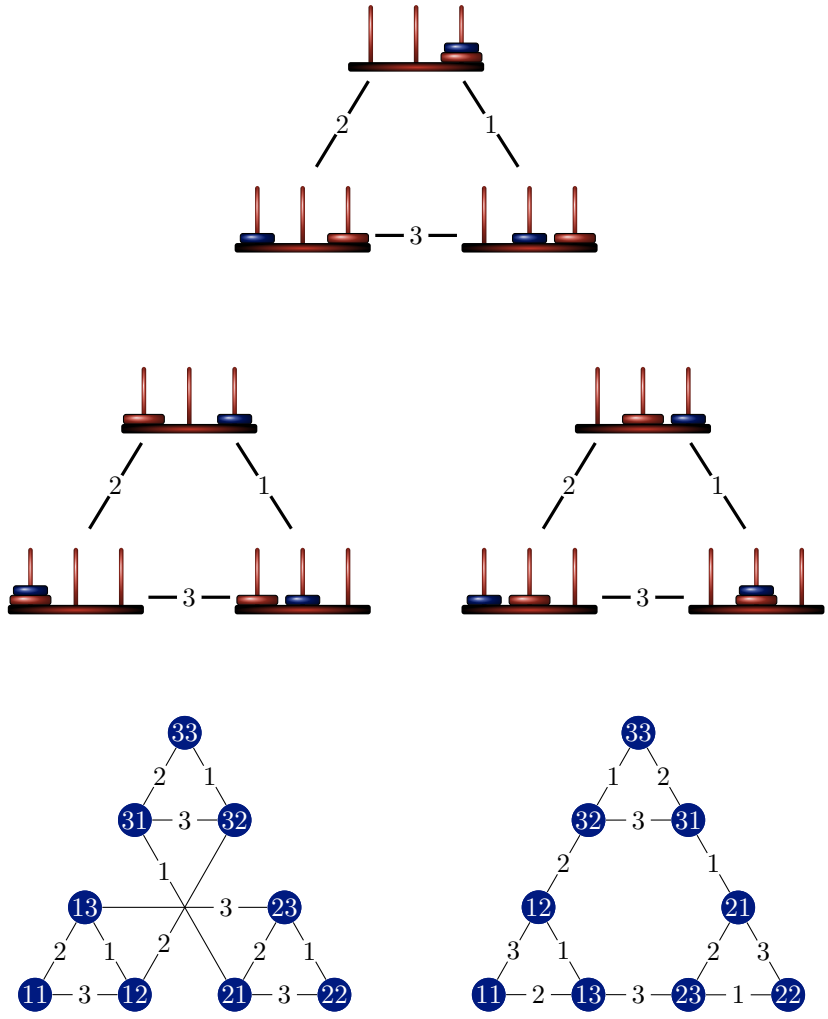


Figure 3.7: Graph of the Tower of Hanoi with two discs. Top: with pictures and only moves of the smallest disc. Bottom both left right: $H(2)$.

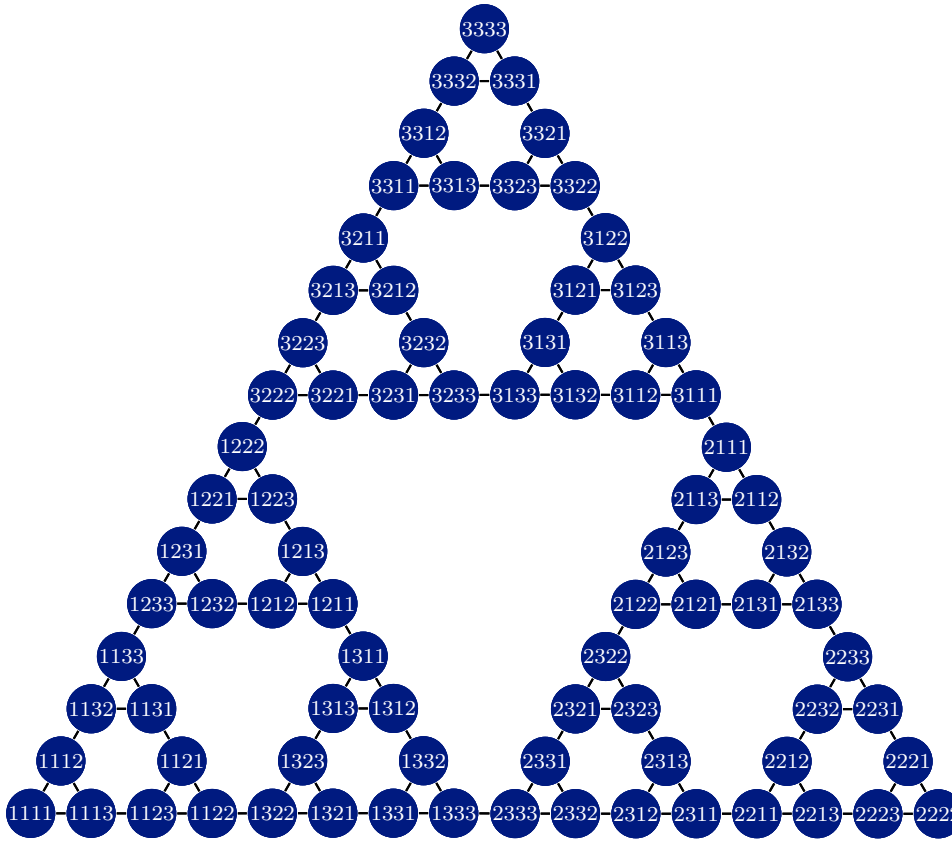


Figure 3.9: $H(4)$

A table of successive positions

For the puzzle with three discs one has:

	2	3	1	2	3	1	2
111	113	123	122	322	321	331	333

This table is constructed from left to right. In the first row the successive moves are given. The second row contains the successive positions obtained by applying the moves in the first row: the red move applied to the red position has the blue position as a result.

Python

What was done by hand can easily be done by computer. Having programmed the three types of moves, every next position of the solution can be printed. The succession of the moves depends only on the parity of the number of discs. We add the function `solution(number)` to the module `hanoi.py`. For the Tower of Hanoi with `number` discs the successive positions are shown on the screen by means of the `print` command.

hanoi.py

```
def solution(number):
    start = number * '1'
    end = number * '3'
    parity = number % 2
    print(start, end = " ")
    if parity == 0:
        peg, position = '3', start
        while position != end:
            position = nextposition(peg, position)
            peg = {'1':'3', '3':'2', '2':'1'}[peg]
            print(position, end = " ")
    else:
        peg, position = '2', start
        while position != end:
            position = nextposition(peg, position)
            peg = {'1':'2', '2':'3', '3':'1'}[peg]
            print(position, end = " ")
```

For $n = 6$ this results in:

```
>>> solution(6)
111111 111112 111132 111133 111233 111231 111221 111222 113222 113223
 113213 113211 113311 113312 113332 113333 123333 123331 123321 12332
2 123122 123123 123113 123111 122111 122112 122132 122133 122233 1222
31 122221 122222 322222 322223 322213 322211 322311 322312 322332 322
333 321333 321331 321321 321322 321122 321123 321113 321111 331111 33
1112 331132 331133 331233 331231 331221 331222 333222 333223 333213 3
33211 333311 333312 333332 333333
```

Exercises

1. How many moves are needed for the solution of the Tower of Hanoi with n discs?
2. Let $W(n)$ be the set of words of length n in the digits 0 and 1. The graph $K(n)$ has $W(n)$ as vertex set. The edges of $K(n)$ are the subsets of $W(n)$ consisting of two words that differ in exactly one place. How many edges does $K(n)$ have?

3. We change the rules of the Tower of Hanoi: moves from peg 1 to peg 3 and visa versa are no longer allowed.
   (i) Make a picture of the corresponding graph.
   (ii) Show that the puzzle can be solved, no matter how many discs are used.
   (iii) How many moves are needed for the puzzle with n discs?

Directions for doing the exercises

a) In exercise 3(ii) it is asked to show something. Formulate the answer in ordinary English.
b) In the exercises 1, 2 and 3 it is asked to determine a number depending on an arbitrary number n. The answer is a formula containing n. Clarify as well as possible the correctness of that formula. Anything helpful, like drawing pictures, is allowed.

Part II

Foundations

Counting one learns at a very young age. This counting is primarily simply the enumeration of numbers: (0,) 1, 2, ... Their names are based on a systematic way of addressing numbers which is nowadays used throughout the world: the decimal system. Chapter 4 is about the basic properties of these numbers, which we will call *natural numbers*. What is minimally required is described by the *axioms of Peano*.
Counting in the sense of determining the *number of elements* of a set is treated in chapter 5. It is described mathematically using *maps* from one set to another. Maps occur everywhere in mathematics.
A *transformation* is a map from a set to itself. Chapter 6 is about properties of repeated application of a transformation.
In chapter 7 the natural numbers are extended with negative numbers. Up to this point no use is made of negative numbers. Together with the natural numbers they form the integral numbers, or *integers* as they are usually called. The way this extension is made makes that the rules of arithmetic for the integers is a direct consequence of these rules for the natural numbers. This method is common practice in all of mathematics.
In chapter 8 the decimal notation is generalized to a notation using an arbitrary base and it is shown how to convert from one notation to another.
Finally in this part fractions are introduced. The integers are extended with fractions by the same method as used for the construction of the integers. For the *rational numbers* thus obtained it is again easily verified that they obey the familiar rules of arithmetic.

4 The Natural Numbers

For every natural number there is a next natural number, its *successor*. We will reduce operations like addition and multiplication to repeatedly taking successors. Thus these operations are fixed, but that does not mean that their properties, the well-known rules of arithmetic, are obvious. These rules require proofs. The starting point for all of this is formed by *Peano's axioms*, which are very basic properties of the natural numbers. The main difficulty lies in the fact that we want to prove general rules that hold for all natural numbers of which there are in fact infinitely many. An important basic property of the natural numbers is the *principle of mathematical induction*. It is a way of dealing with infinity.

Peano's axioms are given in section 4.2. After some first examples of the use of mathematical induction in section 4.3 the operations *addition*, *multiplication* and *exponentiation* will be defined in section 4.4. In the three sections that follow the familiar rules of arithmetic are proved. The way the rules of arithmetic are deduced from Peano's axioms is explained in section 4.5 in which the rules for addition are proved. In the sections that follow the rules for multiplication and exponentiation are proved. The last section is about the *ordering* of natural numbers: the meaning of *less* and *greater* and the properties of these notions.

The rules proved in this chapter are familiar to everybody who has a basic knowledge of arithmetic. Learning how to do proofs in mathematics is often difficult, because in this stage it is not clear what to assume. Here the basic properties are fixed and in this way this difficulty is avoided. A disadvantage of this is that it is not really exciting.

One may just accept the familiar rules for the arithmetic of natural numbers as being obvious and skip Peano's axioms. However, even then it is important to become acquainted with the principle of mathematical induction and with the way notions are defined inductively. Moreover, in this chapter various steps in logical reasoning are made explicit and for many that may be quite instructive.

4.1 Counting

While learning to count one experiences two simple principles:

a) After each number comes another number, a number not encountered before while counting.
b) There are no other numbers than the numbers reached by counting.

In this chapter the meaning of counting is just the enumeration of numbers in the right order. The main use of it is the determination of the number of elements in a set. This is the subject of the next chapter. It is a matter of taste whether one starts counting with 0 or with 1. The numbers we consider here are called *natural numbers*. Since they are used to indicate numbers of elements in a finite set, we let counting start with 0: $\#(\emptyset) = 0$, the empty set has 0 elements. All natural numbers together constitute a set, the set $\mathbb{N}$ of natural numbers, so

$$\mathbb{N} = \{\, n \mid n \text{ is a natural number} \,\} = \{\, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, \dots \,\}.$$

The set of natural numbers $\neq 0$ we denote by $\mathbb{N}^+$. So $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$.

> Starting with 0 is not generally accepted, even not among mathematicians. So in other texts the symbol $\mathbb{N}$ may stand for $\{\, 1, 2, 3, \dots \,\}$. In that case $\mathbb{N}$ is often denoted by $\mathbb{Z}^{\geq 0}$ or $\mathbb{Z}_{\geq 0}$.

For the names of the natural numbers we use words in the ten digits: $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$. Words $\neq 0$ with the left most digit being 0 are excluded. The name of the next natural number (the *successor* of that number) is obtained by replacing the digits 9 at the end of the word (possibly there are none) by digits 0 and by replacing the digit in front of them by the next digit (in the order $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$). For a word of 9's only a 1 is put in front after replacing the 9's by 0's. It is obvious that a number can be retrieved from its successor by the inverse process. The number 0 is the unique number that is not a successor.

> The use of ten digits, i.e. the number ten as a base for our numeral system, is for historic reasons. The term 'numeral system' refers to the way numbers are addressed, their notation. It is a practical choice: we can use our ten fingers for counting. This kind of notation for numbers is very powerful. Many properties of numbers can easily be seen in their notation, for example whether a number is greater than another number.
>
> For a very primitive notation one might use just one digit, a vertical bar | for instance. The notation for 0 then is an empty word and the successor of a number is denoted by placing an extra bar. Then instead of 5 one writes |||||. One can also tally, meaning counting in fives. That is just a bit less primitive and is a first step in the use of five as a base. In chapter 8 numeral systems are considered for any given base.

Python

All programming languages have a data type for natural numbers. However, since we are interested in minimal requirements for operating with natural numbers, the successor function is all we will use when operating with this data type.

All the arithmetic which we will base on the successor function, will be collected in the module `integer.py`. We start with the function `succ` which for every natural number returns its successor.

integer.py

```
def succ(x):
    return x + 1
```

```
>>> succ(5728002869)
5728002870
```

Thus we trust Python that this function is doing what we want. It feels like cheating: we have no idea what what Python actually is doing. Alternatively, we could have taken the description of the successor on page 30: it is a conversion of a word in 0 to 9 into another such word. The reader might do this himself, if so inclined. Or even more basic: represent natural numbers just by strings in | alone, the successor being obtained by adding an extra |.

4.2 Axioms

In Greek Antiquity mathematical thinking made a big leap ahead. In a sense at that time mathematics originated as a discipline. In those times geometry was the principal interest: points, lines, circles. It was based on *primitive notions* (notions which are not defined) and *postulates* (or *axioms*), properties which are not proved. Primitive notions in classical mathematics are for example: point, line, a line passing through a point (or: a point lying on a line). One of the axioms is: through two points there passes a unique line. It is all beautifully written in **Euclid**'s *Elements*, see the nice site

aleph0.clarku.edu/~djoyce/java/elements/elements.html.[1]

Primitive notions are not defined. But they do have properties. Such properties are considered as true if they can be derived from the axioms. In geometry the notions of point and line are not defined, but the axioms attribute to them some properties. If some property can be derived from the axioms there is no need to take it as an axiom. One tries to do mathematics with as few axioms as possible.

[1] The site has many interactive geometric figures based on java. Nowadays for security reasons the use of java in browsers is discouraged.

Euclid ((probably) Alexandria, about 325 BC – Alexandria, about 256 BC)

Euclid was one of the most important mathematicians in Antiquity. Of his life only little is known. For many centuries the series *Elements* (thirteen volumes) have been considered as the base of all mathematics. Until just a few decades ago mathematics education was largely based on Euclid's geometry.

> Modern mathematics is based on sets, as primitive notions are used: set, being element of a set. Thus axiomatic set theory is the foundation of all of mathematics. Here we start from 'naive', nonaxiomatic, set theory. As long as one does not go wild with constructions of sets this will cause no problems. Wanting something like the set of all sets is asking for trouble, see Russell's paradox on page 16.

The Greeks used numbers mainly for counting and numbers also occurred as (ratios of) lengths of line segments and areas of geometrical figures. In Euclid's Elements there are a lot of interesting ideas about numbers, though geometry was the principal interest of the Greeks. They were not inclined to treat numbers the way they treated geometrical objects. The use of numbers was as obvious as the use of logic in mathematics. As with the Romans there was no convenient notation for numbers. Just try to describe in some programming language the successor of a number using the Roman notation for numbers. Not every word in the letters used for numbers in that notation represents a number: it is already complicated to indicate which words actually do. The Italian mathematician **Peano** was the first to describe the system of natural numbers axiomatically.

Peano's Axioms

In Peano's axiomatic treatment of the natural numbers there are three primitive notions: *zero*, *natural number*, the *successor* of a natural number.

The axioms are:

1. Zero is a natural number.
2. The successor of a natural number is a natural number.
3. Zero is not the successor of a natural number.
4. Natural numbers having equal successors are equal.

Giuseppe Peano (Cuneo 1858 – Turin 1932)

Peano was one of the founders of mathematical logic and has contributed to the foundations of mathematics and to the development of a formal logical language.

5. If a property holds for the number zero and if it moreover holds for the successor of every number with that property, then all natural numbers have that property.

Actually, Peano let the natural numbers start with 1 instead of 0, but that hardly matters. The above formulation does not use the terminology of sets. Let's formulate these axioms using the language of sets.

We have:

1. a set $\mathbb{N}$, the elements of which are called natural numbers;
2. an element $\sigma(n) \in \mathbb{N}$, one for each $n \in \mathbb{N}$, called the successor of n;
3. an element $0 \in \mathbb{N}$, called zero.

Peano's axioms:

(N1) There is no $n \in \mathbb{N}$ with $\sigma(n) = 0$.

(N2) For all $m, n \in \mathbb{N}$ with $\sigma(m) = \sigma(n)$ we have $m = n$.

(N3) Let $P(n)$ be a property of natural numbers such that

a) $P(0)$,
b) for all $n \in \mathbb{N}$ with $P(n)$ we also have $P(\sigma(n))$.

Then $P(n)$ for all $n \in \mathbb{N}$.

Axiom (N3) is called the *principle of mathematical induction*. We will use it frequently. Peano's axioms describe our idea of counting. It is clear to us that 0 is not a successor (axiom (N1)). Axiom (N2) is a way to express that a successor of a number is the successor of only that number. Axiom (N3) expresses the idea that there are no other natural numbers than the numbers you can reach by counting, starting from 0: for $P(n)$ take the following property of a natural number n:

If you start counting from 0, then (in principle) you will reach n.

> We will not prove it here, but Peano's axioms are such that they fully determine the system of natural numbers. One says that the axiom system is *categorical*. Starting with 0 and by taking successors repeatedly, an ever growing list of natural numbers is built. It does not contain any repetition and this potentially infinite list contains all natural numbers.

In section 2.3 the relation between properties and subsets was described. Instead of looking in axiom (N3) at a property P, we can also consider the subset U of $\mathbb{N}$ determined by P:

$$U = \{\, n \in \mathbb{N} \mid P(n) \,\}.$$

Then the formulation of (N3) becomes:

(N3') Let U be a subset of $\mathbb{N}$ such that

a) $0 \in U$,
b) for all $n \in U$ also $\sigma(n) \in U$.

Then $U = \mathbb{N}$.

4.3 Reasoning with Numbers

The number 0 is not a successor of a natural number. In fact it is the only natural number with this property. That is not in the axioms. It is a logical consequence of the axioms, or as one says, it can be derived from the axioms. Such a derivation is called a proof. In a proof every assertion (proposition) has to be a direct logical consequence of the axioms and propositions proven beforehand. In mathematics *theorems* are important proven propositions. If such a proposition is less important, then it just called a *proposition*, but of course in that case a proof is still required. The end of a proof is indicated by a □. In other, mostly older, texts 'QED' or 'qed' is used: 'quod erat demonstrandum', Latin for 'what was to be demonstrated'.

4.1 Proposition. *Let m be a natural number different from 0. Then m is the successor of a natural number.*

Proof. We prove that every natural number $\neq 0$ is a successor. Consider the property

$$P(n): \quad n = 0 \ \text{ or } \ n \text{ is a successor.}$$

Clearly $P(0)$ holds.

> Let n be a natural number for which $P(n)$ holds. Because $\sigma(n)$ is a successor, $P(\sigma(n))$ holds as well.

Hence $P(\sigma(n))$ holds for all n for which $P(n)$ holds. From axiom (N3) it follows that $P(n)$ holds for all n. Since $m \neq 0$, it follows that m is a successor. □

This proof is quite formal. If you understand that every natural number can be reached by counting, then the proposition is clearly true.

In order to show in this proof that $P(\sigma(n))$ follows from $P(n)$ for *all* n, we start with the assumption that $P(n)$ holds for a *particular* but otherwise *arbitrary* n. With this fixed n we continue reasoning. As long as the text is indented we reason with this particular n. For this n we derive that also $P(\sigma(n))$ holds. After that we end the indentation and conclude—the n being arbitrary—that for every n satisfying $P(n)$ also $P(\sigma(n))$ holds. Schematically a proof by mathematical induction is along the following lines.

mathematical induction
...
So $P(0)$.
Suppose n is a natural number such that $P(n)$.
...
So $P(\sigma(n))$.
Hence $P(\sigma(n))$ for all $n \in \mathbb{N}$ satisfying $P(n)$.
By mathematical induction $P(n)$ for all $n \in \mathbb{N}$.

The dotted places (...) stand here for reasonings that justify the conclusion that follows.

This scheme for mathematical induction contains a subscheme. Let A be a set. In order to prove a proposition '$Q(a)$ for all $a \in A$', one can take an arbitrary element a of A and prove the property $Q(a)$ (for this arbitrary element a of A).

all
Let $a \in A$.
...
So $Q(a)$.
So $Q(a)$ for all $a \in A$.

Of a we only used the fact that it is an element of A. Therefore, the reasoning that leads to $Q(a)$ is valid for any $a \in A$. In the case of the subscheme of mathematical induction A is the set $\{ n \in \mathbb{N} \mid P(n) \}$ and $Q(n)$ the property $P(\sigma(n))$.

Mathematical induction and the Tower of Hanoi

In the first chapter we concluded that the Tower of Hanoi is solvable for any number of discs. We did so by looking at the graph of the puzzle, and in particular at the relation between the graph and the graph of the puzzle with one extra disc. These graphs tell you more than the solvability of the puzzle alone. For the moment we

concentrate on the solvability. The solvability of the puzzle with n discs can be seen as a property of the natural number n:

$$P(n)\text{: the Tower of Hanoi with } n \text{ discs is solvable.}$$

By mathematical induction we will show that the puzzle is solvable for any number of discs.

Proof. $P(0)$ is trivially true: no moves at all.

> Let n be a natural number satisfying $P(n)$. We will show that then also $P(\sigma(n))$. We assume all $\sigma(n)$ discs being on peg 1 and we will show that by a series of allowed moves they can be transferred to peg 3. Because $P(n)$ holds we can move all discs but the largest to peg 2. Next we move the largest disc. The largest disc is then on peg 3. Since $P(n)$ holds, we can subsequently move the remaining discs from peg 2 to peg 3.

So for each natural number n for which $P(n)$ holds, $P(\sigma(n))$ is holds as well. Because also $P(0)$, we have by mathematical induction that $P(n)$ holds for all $n \in \mathbb{N}$. □

4.4 Addition, Multiplication and Exponentiation

Taking the successor is all we have. Addition is basically done by repeatedly taking successors, multiplication is repeated addition and exponentiation is repeated multiplication. In this section we will make this explicit. It is only about the *definitions* of addition, multiplication and exponentiation; the *properties* of these operations, the rules of arithmetic, are dealt with in the sections 4.5, 4.6 and 4.7.

4.4.1 Addition

When one counts from 5 and stops after the third step, i.e. when one counts $6, 7, 8$, the last number is denoted by $5 + 3$. So $5 + 3 = 8$. We will define generally the meaning of $m + n$ for all natural numbers m and n. To this end we take a fixed, but arbitrary, number $m \in \mathbb{N}$ and describe the meaning of the numbers $m + 0$, $m + 1$, $m + 2$, ... , in this order. Then $m + n$ will be defined for all n. Because we did so for any m, we thus have defined $m + n$ for all m and n.

4.2 Definition. Let m be a natural number. We define $m + n$, the *sum* of m and n (or m *plus* n), for all n by:

$$\begin{cases} m + 0 = m, \\ m + \sigma(n) = \sigma(m + n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

In the first line the meaning of $m + 0$ is given. The second line tells us what $m + \sigma(n)$ means given the meaning of $m + n$. So for example:

$$\begin{aligned} 5 + 0 &= 5 \\ 5 + 1 &= \sigma(5 + 0) = \sigma(5) = 6 \\ 5 + 2 &= \sigma(5 + 1) = \sigma(6) = 7 \\ 5 + 3 &= \sigma(5 + 2) = \sigma(7) = 8 \end{aligned}$$

Since $1 = \sigma(0)$, we have: $m+1 = m+\sigma(0) = \sigma(m+0) = \sigma(m)$. So instead of $\sigma(m)$ we can also write $m + 1$. Then the second line reads $(m + n) + 1 = m + (n + 1)$. The notation $n + 1$ for $\sigma(n)$ we will use frequently.

This definition is an example of an *inductive* or *recursive* definition. The meaning of $m + \sigma(n)$ depends on the meaning of $m + n$. The meaning of $m + 0$ is given directly.

Algorithm

The determination of the sum of m and n goes as follows. Start with the numbers 0 and m and take their successors simultaneously. Repeat this with the resulting two numbers, etc. Continue until the first number is n. Then the second number is $m + n$. This recipe uses the definition of addition in every step. A scheme for $7 + 12$:

0	1	2	3	4	5	6	7	8	9	10	11	12
7	8	9	10	11	12	13	14	15	16	17	18	19

This table is made from left to right: each column determines the next column. The process terminates when the first number in the column is n (here: $n = 12$). This is an example of an *algorithm*: it is a succession of tasks that results in achieving a certain goal. The algorithm terminates since any number will be reached by repeatedly taking successors starting from 0. Such an algorithm can conveniently be done by computer.

Python

We define the sum of natural numbers by repeatedly taking successors. This function `isum` rests on the function `succ` alone.

integer.py

```
def isum(x, y):
    u, v = x, 0
    while v != y: u, v = succ(u), succ(v)
    return u
```

```
>>> isum(255603,16624)
272227
```

The powerful decimal notation of numbers is not used: the successor function is applied 16624 times.

4.4.2 Multiplication

Multiplication is repeated addition:

$$\begin{aligned} 5 \cdot 0 &= 0, \\ 5 \cdot 1 &= 0 + 5 = 5, \\ 5 \cdot 2 &= 5 + 5 = 10, \\ 5 \cdot 3 &= 10 + 5 = 15. \end{aligned}$$

Starting with 0 the number 5 is added 3 times. The meaning of $m \cdot n$ is given by the following inductive definition.

4.3 Definition. Let m be a natural number. We define $m \cdot n$, the *product* of m and n (or m *times* n), for all n by:

$$\begin{cases} m \cdot 0 = 0, \\ m \cdot \sigma(n) = m \cdot n + m \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Writing $n + 1$ for $\sigma(n)$ the second line reads:

$$m \cdot (n + 1) = m \cdot n + m.$$

Algorithm

Start with the numbers 0 and 0. Add m to the second number and replace the first by its successor. Repeat this until the first number is n. Then the second number is $m \cdot n$. A computation of $12 \cdot 11$ using this algorithm.

0	1	2	3	4	5	6	7	8	9	10	11
0	12	24	36	48	60	72	84	96	108	120	132

Python

We add the code for multiplication to the module `integer.py`.

integer.py

```
def iprod(x, y):
    u, v = 0, 0
    while v != y: u, v = isum(u, x), succ(v)
    return u
```

For the computation of the product $15 \cdot 17$ starting with 0 the number 15 is added 17 times:

```
>>> iprod(15,17)
255
```

4.4.3 Exponentiation

Exponentiation is repeated multiplication:

4.4 Definition. Let m be a natural number. For all natural numbers n the meaning of m^n, the n-th *power* of m (or m to the n-th power), is defined by:

$$\begin{cases} m^0 = 1, \\ m^{\sigma(n)} = m^n \cdot m \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

The last line can also be written as: $m^{n+1} = m^n \cdot m$. We have for example:

$$\begin{aligned} 3^0 &= 1 \\ 3^1 &= 1 \cdot 3 = 3 \\ 3^2 &= 3 \cdot 3 = 9 \\ 3^3 &= 9 \cdot 3 = 27 \\ 3^4 &= 27 \cdot 3 = 81 \\ 3^5 &= 81 \cdot 3 = 243 \end{aligned}$$

Notice that $m^0 = 1$ for all m, so in particular $0^0 = 1$. On the other hand $0^m = 0$ for all $m \neq 0$.

Algorithm

Start with the numbers 0 and 1. Replace the second by its product with m and the first by its successor. Repeat this until the first number is n. Then the second is m^n. The computation of 5^8 goes as follows:

0	1	2	3	4	5	6	7	8
1	5	25	125	625	3125	15625	78125	390625

Python

We add to `integer.py` the code for exponentiation.

integer.py

```
def ipower(x, y):
    u, v = 1, 0
    while v != y: u, v = iprod(u, x), succ(v)
    return u
```

For example:

```
>>> ipower(5,8)
390625
```

4.5 Rules for Addition

Addition is a binary operation on natural numbers: to any pair m, n of natural numbers a number $m + n$ is assigned. In general the order of m and n is of importance. For addition however it is not, but that is a rule we still have to prove. It is one of the rules in the following theorem. If you want to call an operation an addition, then you certainly want the first three rules of this theorem to hold.

4.5 Theorem.

- *Addition is* ***associative****:*

  For all $k, m, n \in \mathbb{N}$ $\quad (k + m) + n = k + (m + n)$.

- *The number* 0 *is a* ***neutral element*** *(a* ***zero element****) for the addition:*

  For all $n \in \mathbb{N}$ $\quad n + 0 = 0 + n = n$.

- *Addition is* ***commutative****:*

  For all $m, n \in \mathbb{N}$ $\quad m + n = n + m$.

- ***Cancellation law*** *for the addition:*

  For all $k, m, n \in \mathbb{N}$*:* $\quad$ *if* $k + n = m + n$*, then* $k = m$.

4.6 Definitions. A set together with an associative operation is called a *semigroup*. If there is moreover a neutral element, then it is called a *monoid*. And if the operation is commutative as well, then it is called an *abelian*[2] *monoid*.

So the natural numbers together with the addition is an example of an abelian monoid with a cancellation law .

[2] The adjective 'abelian' refers to Abel, but since it applies to many kinds of mathematical notions—many of them Abel never thought of—it is generally not written with a capital A.

Niels Henrik Abel (Frindø 1802 – Froland 1829)

Many notions in mathematics are named after the Norwegian mathematician Abel. He proved in 1824 that there is no general formula for the solution of equations of degree five, so-called quintic equations, as there is for equations of degree ≤ 4, see also chapter 19. In 2001 the Norwegian government established the Abel price. Since 2002 this prize is yearly awarded and may be considered as the Nobel price for mathematics.

Python

We have defined the function `isum`. This function is based on `succ`. Using the computer one can verify that the addition thus defined is commutative:

```
>>> isum(6369,7087)
13456
>>> isum(7087,6369)
13456
```

With the computer this rule can only be verified for finitely many cases. That is not sufficient for a proof. The instructions differ, but the results are the same.

We will prove the four rules in this theorem separately: the propositions 4.7, 4.8, 4.10 and 4.11. The principle of mathematical induction will be applied repeatedly. We start with associativity.

4.5.1 Associativity

4.7 Proposition. *For all* $k, m, n \in \mathbb{N}$*:* $(k + m) + n = k + (m + n)$.

Proof.

Let k and m be any natural numbers. We will prove the assertion

$$P(n): \quad (k + m) + n = k + (m + n) \quad \text{for all } n \in \mathbb{N}$$

by mathematical induction. First we show that $P(0)$ holds:

$$\begin{aligned} (k + m) + 0 &= k + m && \text{(definition of addition)} \\ &= k + (m + 0) && \text{(definition of addition).} \end{aligned}$$

Assume that n is a natural number for which $P(n)$ holds. Then also $P(n+1)$, because

$$\begin{aligned}(k+m)+(n+1) &= ((k+m)+n)+1 && \text{(definition of addition)}\\ &= (k+(m+n))+1 && \text{(follows from } P(n))\\ &= k+((m+n)+1) && \text{(definition of addition)}\\ &= k+(m+(n+1)) && \text{(definition of addition).}\end{aligned}$$

Hence $P(n+1)$ holds for every n for which $P(n)$ holds. By the principle of mathematical induction we conclude that $P(n)$ holds for all natural numbers n.

So $(k+m)+n = k+(m+n)$ for all $k, m, n \in \mathbb{N}$. □

This proof is quite detailed. It follows the **all** scheme and within this scheme the **mathematical induction** scheme. Usually one writes shorter proofs:

all (short)
Let $a \in A$.
...
So $Q(a)$.

Thus omitting the obvious conclusion. This makes a proof by mathematical induction less elaborate:

mathematical induction (short)
...
So $P(0)$.
Let n be a natural number with $P(n)$.
...
Hence $P(n+1)$.
By mathematical induction $P(n)$ for all $n \in \mathbb{N}$.

In $k+m+n$ parentheses may be placed in two ways. Associativity means that the way they are placed is irrelevant for the meaning of the expression. This not only holds for a sum with three terms but also for sums with more terms. For example with four terms:

$$\begin{aligned}k+((l+m)+n) = ((k+l)+m)+n &= (k+l)+(m+n)\\ = k+(l+(m+n)) &= (k+(l+m))+n.\end{aligned}$$

Since the way parentheses are placed makes no difference for the meaning, we rather do not place them at all. We simply agree that $k+m+n$ has the same meaning as this expression with any placement of parentheses.

4.5.2 Neutrality of 0

4.8 Proposition. *For all* $n \in \mathbb{N}$*:* $n + 0 = 0 + n = n$.

Proof. By the definition of addition $n + 0 = n$ for all $n \in \mathbb{N}$. We will prove by mathematical induction that

$$Q(n): \quad 0 + n = n$$

holds for all $n \in \mathbb{N}$. The proposition $Q(0)$ follows from the definition of the addition.

> Suppose n is a natural number such that $Q(n)$, that is $0 + n = n$. Then $0 + n + 1 = n + 1$. So $Q(n + 1)$ holds.

By mathematical induction it follows that $Q(n)$ holds for all $n \in \mathbb{N}$. □

4.5.3 Commutativity

Before proving the commutativity of the addition we first derive a special case.

4.9 Lemma. *For all* $m \in \mathbb{N}$*:* $m + 1 = 1 + m$.

Proof. We will prove by mathematical induction that the following proposition holds for all $m \in \mathbb{N}$:

$$R(m): \quad m + 1 = 1 + m.$$

$R(0)$ follows from the definition of the addition: $0 + 1 = 1 = 1 + 0$.

> Let m be a natural number such that $R(m)$, that is $m + 1 = 1 + m$. Then $m + 1 + 1 = 1 + m + 1$. So $R(m + 1)$ holds also.

By mathematical induction it follows that $R(m)$ holds for all $m \in \mathbb{N}$. □

4.10 Proposition. *For all* $m, n \in \mathbb{N}$*:* $m + n = n + m$.

Proof. Let m be any natural number. We will prove by mathematical induction that the following proposition holds for all $n \in \mathbb{N}$:

$$S(n): \quad m + n = n + m.$$

From proposition 4.8 follows $0 + m = m = m + 0$, so $S(0)$ holds.

> Let n be a natural number such that $S(n)$, that is $m + n = n + m$. Then by lemma 4.9:
>
> $$m + n + 1 = n + m + 1 = n + 1 + m.$$
>
> Hence also $S(n + 1)$.

By mathematical induction it follows that $S(n)$ for all $n \in \mathbb{N}$. □

Associativity of the addition made it unnecessary to place parentheses in a sum with more then two terms. Since 0 is a neutral element terms 0 may be omitted. The commutativity of the addition implies that also the order of the terms makes no difference for the meaning of the expression.

4.5.4 Cancellation

4.11 Proposition. *For all $k, m, n \in \mathbb{N}$: if $k + n = m + n$, then $k = m$.*

Proof. Let k and m be any natural numbers. We will prove by mathematical induction that the proposition

$$T(n): \quad \text{if } k + n = m + n, \text{ then } k = m$$

holds for all natural numbers n. For $n = 0$ this follows directly from the definition of the addition.

Suppose n is a natural number such that $T(n)$.

Assume that $k + n + 1 = m + n + 1$. Because the successors of $k + n$ and $m + n$ coincide, $k + n$ and $m + n$ are equal as well (axiom). From $T(n)$ it follows that $k = m$.

Hence $T(n + 1)$.

By mathematical induction it follows that $T(n)$ for all $n \in \mathbb{N}$. □

In order to prove a proposition of the form

if P, then Q

one can assume P and prove Q using this assumption. Scheme:

if, then
Suppose P
...
Hence Q.
So: if P, then Q.

4.5.5 Extra Rules

We will need two extra rules for the addition of natural numbers.

4.12 Proposition. *Let m and n be natural numbers such that $m + n = 0$. Then $m = n = 0$.*

Proof.

Suppose $n \neq 0$. From proposition 4.1 it follows that n is a successor: $n = n' + 1$ for an $n' \in \mathbb{N}$. Then $m + n = m + n' + 1$. Hence $m + n$ is a successor and therefore $m + n \neq 0$ (axiom). Contradiction.

So $n = 0$. Then also $m = 0$. □

This is an example of a *proof by contradiction*. A way to prove a proposition of the form 'not P' is given in the following scheme.

not
Suppose P.
...
Contradiction.
Hence not P.

A proof by contradiction is based on the idea that a proposition is either true or false:

contradiction
Suppose not P.
...
Contradiction.
Hence P.

4.13 Proposition. *Let m and n be natural numbers such that $m + n = 1$. Then $m = 0$ or $n = 0$.*

Proof.

Suppose $n \neq 0$. Then n is a successor: $n = n' + 1$ for an $n' \in \mathbb{N}$ (proposition 4.1). We have $m + n' + 1 = 1 = 0 + 1$. The numbers $m + n'$ and 0 have the same successor. So $m + n' = 0$. From proposition 4.12 it follows that $m = 0$.

Hence $m = 0$ or $n = 0$. □

In order to prove a proposition of the form 'P or Q' it suffices to prove P:

or
...
So P.
Hence, P or Q.

Mostly, as above, for proving a proposition of the form 'P or Q' another approach is more effective, in particular when it is not clear whether P holds or Q holds:

if not, then
Suppose not P.
...
So Q
Hence P or Q.

For the sake of completeness we also include a scheme for a proposition of the the form 'P and Q'. We have a proof if we have proofs for both P and Q.

and
...
So P.
...
So Q.
Hence P and Q.

4.6 Rules for Multiplication

As there are rules for addition, there are rules for multiplication. One of these rules connects addition and multiplication. That is no wonder, since multiplication is repeated addition.

4.14 Theorem.

- *Multiplication is* ***associative****:*

  For all $k, m, n \in \mathbb{N}$ $\quad (k \cdot m) \cdot n = k \cdot (m \cdot n)$.

- *The number* 1 *is a* ***neutral element*** *(or* ***unit element****) for the multiplication:*

  For all $n \in \mathbb{N}$ $\quad n \cdot 1 = 1 \cdot n = n$.

- *Multiplication is* ***commutative****:*

  For all $m, n \in \mathbb{N}$ $\quad m \cdot n = n \cdot m$.

- ***Cancellation law*** *for the multiplication:*

  For all $k, m, n \in \mathbb{N}$*:* $\quad$ *if* $m \cdot k = n \cdot k$ *and* $k \neq 0$*, then* $m = n$.

- *Multiplication is* ***distributive*** *over addition:*

  For all $k, m, n \in \mathbb{N}$ $\quad k \cdot (m + n) = k \cdot m + k \cdot n$.

The set $\mathbb{N}$ together with the multiplication is an abelian monoid. From lemma 4.21 it follows that multiplication is an operation in $\mathbb{N}^+$ as well. The set $\mathbb{N}^+$ together with the multiplication is an abelian monoid with cancellation law.

Below the rules will be proved separately: the propositions 4.20, 4.15, 4.18, 4.19 and 4.22.

4.6.1 Neutrality of 1

4.15 Proposition. *For all natural numbers n:* $n \cdot 1 = 1 \cdot n = n$.

Proof. First we prove that $n \cdot 1 = n$ for any $n \in \mathbb{N}$:

$$\begin{aligned} n \cdot 1 &= n \cdot 0 + n && \text{(definition of multiplication)} \\ &= 0 + n && \text{(definition of multiplication)} \\ &= n && \text{(proposition 4.8).} \end{aligned}$$

We will prove $1 \cdot n = n$ for all natural numbers n by mathematical induction. For $n = 0$ it follows from the definition of multiplication.

Suppose n is a natural number such that $1 \cdot n = n$. Then

$$\begin{aligned} 1 \cdot (n+1) &= 1 \cdot n + 1 && \text{(definition of multiplication)} \\ &= n + 1. \end{aligned}$$

The proposition now follows by mathematical induction. □

4.6.2 Commutativity

4.16 Lemma. *For all natural numbers n:* $0 \cdot n = 0$.

Proof. We prove by mathematical induction that the proposition

$$P(n): \quad 0 \cdot n = 0$$

holds for all natural numbers n. For $n = 0$ this follows from the definition of multiplication.

Suppose n is a natural number such that $P(n)$. Then

$$\begin{aligned} 0 \cdot (n+1) &= 0 \cdot n + 0 && \text{(definition of multiplication)} \\ &= 0 + 0 && (P(n)) \\ &= 0 && \text{(definition of addition).} \end{aligned}$$

Hence $P(n+1)$.

By mathematical induction it follows that $P(n)$ for all $n \in \mathbb{N}$. □

4.17 Lemma. *For all natural numbers m and n:* $(n+1)m = nm + m$.

Proof. Let n be any natural number. We prove by mathematical induction that

$$Q(m): \quad (n+1)m = nm + m$$

holds for all natural numbers m. From the definitions of addition and multiplication it follows that $Q(0)$ holds.

Let m be a natural number such that $Q(m)$. Then

$$\begin{aligned}(n+1)(m+1) &= (n+1)m + n + 1 && \text{(definition of multiplication)}\\ &= nm + m + n + 1 && (Q(m))\\ &= nm + n + m + 1 && \text{(commutativity of addition)}\\ &= n(m+1) + m + 1 && \text{(definition of multiplication).}\end{aligned}$$

It follows by mathematical induction that $Q(m)$ for all natural numbers m. □

4.18 Proposition. *For all $m, n \in \mathbb{N}$: $mn = nm$.*

Proof. Let m be any natural number. We prove by mathematical induction the proposition

$R(n)$: $mn = nm$.

$R(0)$ follows from the definition of multiplication and lemma 4.16.

Suppose n is a natural number such that $R(n)$. Then

$$\begin{aligned}m(n+1) &= mn + m && \text{(definition of multiplication)}\\ &= nm + m && (R(n))\\ &= (n+1)m && \text{(lemma 4.17).}\end{aligned}$$

Hence $R(n+1)$.

It follows by mathematical induction that $R(n)$ for all $n \in \mathbb{N}$. □

4.6.3 Distributivity

4.19 Proposition. *For all $k, m, n \in \mathbb{N}$: $k(m+n) = km + kn$.*

Proof. Let k and m be any natural numbers. We prove the proposition

$S(n)$: $k(m+n) = km + kn$

by mathematical induction. From the definitions of addition and multiplication follow $k(m+0) = km$ and $km + k0 = km + 0 = km$, so $S(0)$.

Suppose n is a natural number such that $S(n)$. Then

$$\begin{aligned}k(m+n+1) &= k(m+n) + k && \text{(definition of multiplication)}\\ &= km + kn + k && (S(n))\\ &= km + k(n+1) && \text{(definition of multiplication).}\end{aligned}$$

Hence $S(n+1)$.

By mathematical induction it follows that $S(n)$ holds for all $n \in \mathbb{N}$. □

4.6.4 Associativity

4.20 Proposition. *For all $k, m, n \in \mathbb{N}$:* $(km)n = k(mn)$.

Proof. Let k and m be any natural numbers. We prove by mathematical induction the proposition

$$T(n): \quad (km)n = k(mn).$$

$T(0)$ follows from the definition of multiplication.

Let n a natural number such that $T(n)$. Then

$$\begin{aligned}
(km)(n+1) &= (km)n + km && \text{(definition of multiplication)} \\
&= k(mn) + km && (T(n)) \\
&= k(mn + m) && \text{(distributivity)} \\
&= k(m(n+1)) && \text{(definition of multiplication).}
\end{aligned}$$

By mathematical induction it follows that $T(n)$ holds for all natural numbers n. □

Because multiplication is associative and commutative and since 1 is a neutral element (equivalently, $\mathbb{N}$ together with the multiplication is an abelian monoid), in products with more than two factors parentheses may be omitted, the order of the factors makes no difference and factors 1 can be left out.

4.6.5 Cancellation

4.21 Lemma. *Let m and n be natural numbers with $mn = 0$. Then $m = 0$ or $n = 0$.*

Proof.

Suppose $m \neq 0$ and $n \neq 0$. Then m and n are successors: $m = m' + 1$ and $n = n' + 1$ for natural numbers m' and n'. Then mn is a successor as well: $mn = (m' + 1)(n' + 1) = m'n' + n' + m' + 1$. Contradiction.

Hence $m = 0$ or $n = 0$. □

4.22 Proposition. *Let n be a natural number $\neq 0$. Then for all $k, m \in \mathbb{N}$:*

$$\text{if } kn = mn, \text{ then } k = m.$$

Proof. We prove by mathematical induction the proposition

$U(m)$: for all $k \in \mathbb{N}$: if $kn = mn$, then $k = m$.

Suppose $k \in \mathbb{N}$ with $kn = 0 \cdot n$, or equivalently $kn = 0$. Then, since $n \neq 0$, from lemma 4.21 it follows that $k = 0$.

So $U(0)$.

Let m be a natural number with $U(m)$.

Suppose $k \in \mathbb{N}$ such that $kn = (m+1)n$. Then $kn \neq 0$, since $m+1 \neq 0$ and $n \neq 0$. So also $k \neq 0$. This means that k is a successor: $k = k' + 1$ for a $k' \in \mathbb{N}$. Then $(k'+1)n = (m+1)n$ and so $k'n + n = mn + n$. From the cancellation law for the addition it follows that $k'n = mn$. From $U(m)$ follows that $k' = m$, that is $k = m + 1$.

Hence $U(m+1)$.

By mathematical induction it follows that $U(m)$ holds for all $m \in \mathbb{N}$. □

An extra rule for the multiplication of natural numbers:

4.23 Proposition. *Let m and n be natural numbers with $mn = 1$. Then $m = n = 1$.*

Proof. Because $mn \neq 0$ we have $m \neq 0$. So m is a successor: $m = m' + 1$ for an $m' \in \mathbb{N}$. Then $m'n + n = 1$. Since also $n \neq 0$ it follows that $m'n = 0$ (lemma 4.13). So $m' = 0$, that is $m = 1$. Then $n = 1$ as well. □

4.7 Rules for Exponentiation

4.24 Proposition. *For all $k, m, n \in \mathbb{N}$: $k^m k^n = k^{m+n}$.*

Proof. Let k and m be any natural numbers. We prove the proposition

$P(n)$: $k^m k^n = k^{m+n}$

by mathematical induction. $P(0)$ follows from $k^m k^0 = k^m 1 = k^m$ and $k^{m+0} = k^m$.

Let n be a natural number such that $P(n)$. We have

$$\begin{aligned} k^m k^{n+1} &= k^m k^n k && \text{(definition of exponentiation)} \\ &= k^{m+n} k && (P(n)) \\ &= k^{m+n+1} && \text{(definition of exponentiation).} \end{aligned}$$

Hence $P(n+1)$.

By mathematical induction it follows that $P(n)$ holds for all $n \in \mathbb{N}$. □

4.25 Proposition. *For all $k, m, n \in \mathbb{N}$: $(k^m)^n = k^{mn}$.*

Proof. Let k and m be any natural numbers. We prove the proposition

$$Q(n): \quad (k^m)^n = k^{mn}$$

by mathematical induction. $Q(0)$ follows from $(k^m)^0 = 1$ and $k^{m0} = k^0 = 1$.

Let n be a natural number such that $Q(n)$. We have

$$\begin{aligned}
(k^m)^{n+1} &= (k^m)^n k^m && \text{(definition of exponentiation)} \\
&= k^{mn} k^m && (Q(n)) \\
&= k^{mn+m} && \text{(proposition 4.24)} \\
&= k^{m(n+1)} && \text{(definition of multiplication).}
\end{aligned}$$

So $Q(n+1)$.

By mathematical induction it follows that $Q(n)$ holds for all n. □

4.26 Proposition. *For all $k, m, n \in \mathbb{N}$:* $(km)^n = k^n m^n$.

Proof. Let k and m be any natural numbers. We prove the proposition

$$R(n): \quad (km)^n = k^n m^n$$

by mathematical induction. $R(0)$ follows from $(km)^0 = 1$ and $k^0 m^0 = 1 \cdot 1 = 1$.

Let n be a natural number such that $R(n)$. We have

$$\begin{aligned}
(km)^{n+1} &= (km)^n km && \text{(definition of exponentiation)} \\
&= k^n m^n km && (R(n)) \\
&= k^n k m^n m && \text{(commutativity of multiplication)} \\
&= k^{n+1} m^{n+1} && \text{(definition of exponentiation).}
\end{aligned}$$

Hence $R(n+1)$.

By mathematical induction it follows that $R(n)$ holds for all n. □

4.8 Ordering

4.27 Definition. Let m and n be natural numbers. Let x be a natural number such that $m + x = n$. Then x is called the *difference* of n and m. Notation: $x = n - m$.

> Note that if such an x exists, there is no other: if y is a natural number with $m + y = n$, then $m + y = m + x$, and so by the cancellation law for addition: $y = x$. That is why we can speak of *the* difference.

The difference of n and m exists if one arrives at n when starting to count from m. If in this process the successor is taken x times, the difference is x.

4.28 Definition. Let m and n be natural numbers. We say that m *is less than or equal to* n if there exists an x such that $m + x = n$. Notation: $m \leq n$ (or $n \geq m$ and then we say that n is *greater than or equal to* m). If $x \neq 0$, so if $m \neq n$, then we say that m is *less than* n. Notation: $m < n$ (or $n > m$: n *greater than* m).

So: $m \leq n \iff$ the difference of n and m exists.
$m < n \iff$ the difference of n and m exists and is not 0.

4.29 Proposition. *The relation* $\leq$ *is an ordering of* $\mathbb{N}$*, i.e.:*

(i) $n \leq n$ *for all* $n \in \mathbb{N}$,
(ii) *for all* $k, m, n \in \mathbb{N}$: *if* $k \leq m$ *and* $m \leq n$*, then* $k \leq n$,
(iii) *for all* $m, n \in \mathbb{N}$: *if* $m \leq n$ *and* $n \leq m$*, then* $m = n$.

Proof.

(i) Since $n + 0 = n$, we have $n \leq n$.
(ii) If $k + x = m$ and $m + y = n$, then $k + x + y = m + y = n$ and so $k \leq n$.
(iii) If $m + x = n$ and $n + y = m$, then $m + x + y = m = m + 0$. From the cancellation law it follows that $x + y = 0$. So $x = 0$ (lemma 4.12), that is $m = n$. □

There are no natural numbers between a natural number and its successor:

4.30 Lemma. *Let* n *and* k *be natural numbers with* $n \leq k \leq n + 1$*. Then* $k = n$ *or* $k = n + 1$.

Proof. There are $x, y \in \mathbb{N}$ such that $n + x = k$ and $k + y = n + 1$. Then $n + x + y = k + y = n + 1$. From the cancellation law for the addition it follows that $x + y = 1$. By proposition 4.13 we have $x = 0$ or $y = 0$, so $n = k$ or $k = n + 1$. □

The next proposition states that the natural numbers are *totally* ordered by $\leq$: for natural numbers m and n we have $m \leq n$ or $n \leq m$. Both $m \leq n$ and $n \leq m$ only if $m = n$. So exactly one of the following three propositions is true: $m < n$, $m = n$, $n < m$.

4.31 Proposition. *Let* m *and* n *be natural numbers. Then* $m \leq n$ *or* $n \leq m$.

Proof. Let m be any natural number. We prove by mathematical induction that the proposition

$P(n)$: $m \leq n$ or $n \leq m$

holds for all natural numbers n. Clearly $P(0)$ holds: $0 \leq m$.

Let n be a natural number such that $P(n)$. If $m \leq n$, then $m \leq n + 1$. So suppose that $m \leq n$ is not the case. By $P(n)$ we have $n < m$ and so $n + x = m$ for an $x \in \mathbb{N}^+$. Since $x \neq 0$, the number x is a successor: $x = y + 1$, with $y \in \mathbb{N}$. Then $n + 1 + y = m$ and so $n + 1 \leq m$. □

The next proposition describes a connection with addition.

4.32 Proposition. *Let m, n and k be natural numbers. Then*

$$m + k \leq n + k \iff m \leq n.$$

Proof.

$\Leftarrow$: Suppose $m \leq n$. Then there is an $x \in \mathbb{N}$ such $n = m + x$. Then also $m + k + x = n + k$ and so $m + k \leq n + k$.

$\Rightarrow$: Suppose $m + k \leq n + k$. Then there is an $x \in \mathbb{N}$ such that $n + k = m + k + x$. From the cancellation law for the addition it follows that $n = m + x$. Hence $m \leq n$. □

A connection with multiplication is described in the next proposition.

4.33 Proposition. *Let m, n and k be natural numbers. Then:*

(i) *if $m \leq n$, then $mk \leq nk$;*
(ii) *if $mk \leq nk$ and $k \neq 0$, then $m \leq n$.*

Proof.

(i) Suppose $m + x = n$ for an $x \in \mathbb{N}$. Then $mk + xk = nk$ and so $mk \leq nk$.
(ii) Suppose $mk \leq nk$ and $k \neq 0$.

> Suppose that not $m \leq n$. Then by proposition 4.31 $n < m$, that is $n + x = m$ for an $x \in \mathbb{N}^+$. Then $nk + xk = mk$ and so $nk > mk$, because $xk \neq 0$. Contradiction.

Hence $m \leq n$. □

Since for any pair of natural numbers m, n we have $m \leq n$ or $n \leq m$, one of the differences $n - m$ and $m - n$ exists.

4.34 Definitions and notations. If the equation $m \cdot x = n$ is solvable, then we say that n is a *multiple* of m and if $m \neq 0$ we denote the solution as $\frac{n}{m}$. (By the cancellation law for the multiplication this solution is unique.)

If the equation $x^m = n$ is solvable, then we say that n is an *m-th power* and if $m \neq 0$ we denote the solution as $\sqrt[m]{n}$. This solution is unique if $x \neq 0$. (This follows by mathematical induction using the cancellation law for the multiplication.)

If the equation $m^x = n$ is solvable, then we say that n is a *power* of m and if $m \geq 2$ we denote the solution as ${}^m\log n$. Also this one is unique.

An important goal of the extensions of the number system is to have solutions for equations, but that is not all we want. We also insist the normal rules for arithmetic to remain valid. Maybe that is asking too much. Anyway, the existence of such extensions is far from obvious.

Algorithm

Proposition 4.31 makes it so that there is an algorithm for determining the difference of two natural numbers if it exists: take successors of both numbers, repeat this for the result, etc. Do so until one of the original numbers reappears. The proposition states that this will happen. In a scheme for the numbers 6 and 19:

0	1	2	3	4	5	6	7	8	9	10	11	12	13
6	7	8	9	10	11	12	13	14	15	16	17	18	19
19	20	21	22	23	24	25	26	27	28	29	30	31	32

This shows that $6 \leq 19$ and moreover, that $19 - 6 = 13$.

Python

The algorithm can easily be converted to Python-code. We add it to `integer.py`.

integer.py

```
def idiff(x, y):
    u, v, w = 0, x, y
    while v != y and w != x: u, v, w = succ(u), succ(v), succ(w)
    return u,v,w

def leq(x, y):
    return idiff(x, y)[1] == y

def geq(x, y):
    return idiff(x, y)[2] == x

def difference(x, y):
    return idiff(x, y)[0]
```

Then:

```
>>> leq(567,120)
False
>>> geq(567,120)
True
>>> difference(567,120)
447
```

Exercises

1. In exercise 3 of chapter 1 we saw that the Tower of Hanoi remains solvable if moves between pegs 1 and 3 are not allowed, whatever the number of discs. We observed

this by looking at the corresponding graph. Show by mathematical induction that this puzzle is solvable for any number of discs.

2. Let a and b be natural numbers. Then
$$(a+b)^2 = a^2 + 2ab + b^2.$$
Prove this.

3. Natural numbers occur as numbers of elements of sets. Having 3 pairwise disjoint sets of 5 elements, the 15 elements can also be grouped in 5 sets of 3 elements, see Figure 4.1. The number of elements of a set can be determined by counting

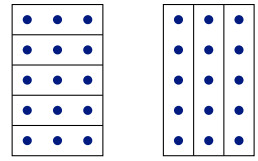


Figure 4.1: Commutativity of the multiplication

the elements. It is crucial that the result is independent of the order in which the elements are counted. Rules of arithmetic—such as here the commutativity of the multiplication—can be visualized by grouping the elements in different ways. Visualize the distributivity of the multiplication over the addition of natural numbers.

4. Let k and m be natural numbers such that $k \leq m$.
   (i) Prove that $k^2 \leq m^2$.
   (ii) Prove that $k^n \leq m^n$ for all $n \in \mathbb{N}$.

5. Let k, m and n be natural numbers such that $m \leq n$ and $k \neq 0$. Prove that $k^m \leq k^n$.

6. (i) Show that exponentiation is not associative. (The convention is that $k^{(m^n)}$ can be denoted as k^{m^n}.)
   (ii) We want to use the notation $m * n$ for repeated exponentiation: $m * 1 = m$, $m * 2 = m^m$, $m * 3 = m^{m^m}$, ... How should $m * n$ be defined? What would $m * 0$ be?

7. Prove that for any natural number n the number $n^5 - n$ is a multiple of 5.

8. Prove that for all natural numbers n we have: $n < 2^n$.

9. Let a and b be natural numbers with $b \leq a$. Prove the identity
$$(a-b)^2 = (a^2 + b^2) - 2ab.$$

10. Let a and b be natural numbers such that $b \leq a$. Prove the identity
$$(a-b)(a+b) = a^2 - b^2.$$

11. Prove that for all $n \in \mathbb{N}$ the inequality $2^{n+1} \geq n^2 + n + 2$.

Directions for doing the exercises

a) Indicate exactly in the exercises 2, 4, 5, 7, 8, 9 and 10 which propositions or rules are used.
b) In the exercises 7, 9 and 10 the difference of natural numbers occurs. The number $a - b$ is, if it exists, the number that added to b gives a. Use that. Since negative numbers are not yet introduced in this stage $a - b$ can not be seen as $a + (-b)$.

5 Counting

Natural numbers are made for counting - counting in the sense of determining the number of elements of a set. Intuitively it is clear what is meant by 'number of elements', we have been using this in Part I. In this chapter a more precise meaning of counting is introduced. As a consequence we have to prove properties of counting which are intuitively obvious, e.g. the outcome being independent of the order in which elements are counted. As is to be expected, in these proofs mathematical induction is used frequently; it is a fundamental property of the natural number system. The operations addition, multiplication and exponentiation of natural numbers are closely related to operations on sets. We will make this explicit. The counting in this chapter is very elementary. Chapter 11 deals with a smarter kind of counting, also known as combinatorics. In section 5.7 elementary, but important, counting principles are treated.

Maps and all related notions, as composition of maps, injectivity and such are frequently used in mathematics. In this book they will be used regularly. In this chapter the emphasis is on their connection with counting.

5.1 Maps

For the comparison of sets we use maps between these sets. First we define this very important notion in mathematics of map.

5.1 Definition. A *map* (or *mapping*) f from a set A to a set B consists of

a) a set A, the *domain* of f,
b) a set B, the *codomain* of f,
c) for each element a of A an element $f(a)$ of B, the *image* of a under f.

5.2 Notations. If f is a map from A to B, then we denote this as $f\colon A \to B$ or as $A \xrightarrow{f} B$. To indicate that a $b \in B$ is the image under f of an $a \in A$, so $f(a) = b$, we also write $f\colon a \mapsto b$. Note the difference in the use of the symbols $\to$ and $\mapsto$.

5.3 Terminology. Another word for map is *function*. A map from A to B is also called a function on A with values in B. In the function terminology one usually calls $f(a)$ the *value* of the function f in a. Here we will mainly use the word function if the codomain is a set of numbers and moreover, the codomain is not

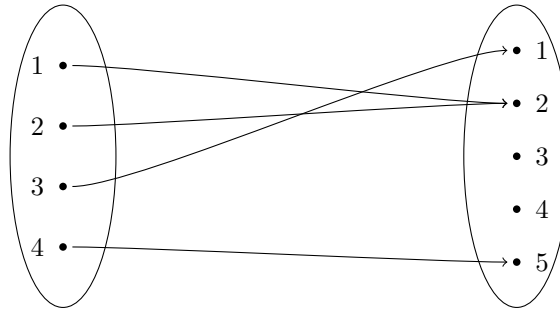


Figure 5.1: Picture of a map

seen as an important aspect of the structure. We then simply speak of a function on A. The function values are numbers. Extension of the number system leads to having more functions on A. For the time being we only have functions with values in the natural numbers.

5.4 Example. Let $A = \{1, 2, 3, 4\}$ and $B = \{1, 2, 3, 4, 5\}$. A map f from A to B is determined by the assignment of an element of B to each $a \in A$; that element is denoted by $f(a)$. So for example: $f(1) = 2$, $f(2) = 2$, $f(3) = 1$ and $f(4) = 5$. In this way we have an image $f(a) \in B$ for each $a \in A$. See Figure 5.1 for a picture of this map. This map may be denoted by $f = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 1 & 5 \end{smallmatrix}\right)$. The top row is an enumeration of the elements of A and for each of these elements its image is given directly underneath. The notation reveals the domain, but not completely the codomain, only the image of elements in the codomain.

> We will not often use this way to denote a map $f\colon A \to B$. It is not a generally accepted notation. In many cases there are better ways to denote a map.

5.5 Examples. For each natural number n there is its successor $\sigma(n)$. This can be seen as a map σ from $\mathbb{N}$ to $\mathbb{N}$, in fact a transformation of $\mathbb{N}$:

$$\sigma\colon \mathbb{N} \to \mathbb{N}, \quad n \mapsto \sigma(n).$$

Other examples of transformations of $\mathbb{N}$ are:

$$\begin{aligned} &\alpha_m\colon \mathbb{N} \to \mathbb{N}, \quad n \mapsto n + m && \text{(add } m\text{)}, \\ &\mu_m\colon \mathbb{N} \to \mathbb{N}, \quad n \mapsto nm && \text{(multiply with } m\text{)}. \end{aligned}$$

> The maps α_m and μ_m are given by a formula, namely $n + m$, respectively nm. Or, if you prefer an x in a formula, $x + m$ and xm. In secondary school mathematics a map (a function) is often defined as a formula. Note that in the definition given here there is no mention of a formula, see also example 5.4, where no formula was given. All that is required, is designating to every element of the domain an image element in the codomain.

René Descartes (La Haye (now Descartes) 1596 – Stockholm 1650)

Descartes, or Cartesius, made the application of algebra to geometry possible by the introduction of coordinates, nowadays often called Cartesian coordinates. Long before Descartes, already in the 14th century, coordinates were used by the French mathematician N. Oresme. He lived in The Netherlands during a long part of his lifetime.

5.6 Definition. Let $f\colon A \to B$ and $U \subseteq A$. The *restriction* of f to U is the map $U \to B$, $u \mapsto f(u)$. A map $g\colon A' \to B$, where $A' \supseteq A$ is called a *prolongation* of f to A' if f is the restriction of g to A .

A map has a unique restriction to a given subset. There is no unique prolongation of a map $f\colon A \to B$ to an $A' \supseteq A$, unless $A' = A$ or $\#(B) = 1$.

5.2 The Graph of a Map

An *ordered pair* (a, b) has a first element a and a second element b. We allow that $a = b$. Characteristic for ordered pairs is that they satisfy

$$(a, b) = (c, d) \iff a = c \text{ and } b = d.$$

One way to see an ordered pair as a set is to define it as follows:

$$(a, b) = \{\{a\}, \{a, b\}\}.$$

See also exercise 1.

5.7 Definition. Let A and B be sets. The *Cartesian product* of A and B is the set

$$A \times B = \{\, (a, b) \mid a \in A \text{ and } b \in B \,\}.$$

The Cartesian product is also called the *product* for short.

A map $f\colon A \to B$ consists of a domain A, a codomain B and for each $a \in A$ an $f(a) \in B$. The map f may also be given by the subset of $A \times B$ consisting of all ordered pairs with an element of the domain A as the first element and its image in B as the second.

5.8 Definition. The *graph* of a map $f\colon A \to B$ is the following subset of $A \times B$:

$$\Gamma(f) = \{\, (a, f(a)) \mid a \in A \,\}.$$

Note that the word graph has now two meanings: the graph of a map and the graph as a structure consisting of vertices and nodes as defined earlier. When used the meaning should be clear from the context.

5.9 Example. The graph of the map $f\colon \{1,2,3,4\} \to \{1,2,3,4,5\}$ of example 5.4 is a subset of the product $\{1,2,3,4\} \times \{1,2,3,4,5\}$. See Figure 5.2. The dots correspond to elements of the Cartesian product. The graph is the subset of the elements indicated by the big dots. Note that in such a picture every vertical line contains exactly one element of the graph.

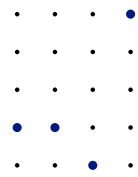


Figure 5.2: The graph of the map f of example 5.4

5.10 Examples. The graphs of σ, α_m and μ_m of examples 5.5 are subsets of $\mathbb{N} \times \mathbb{N}$. Pictures of these graphs are of course incomplete, see Figure 5.3.

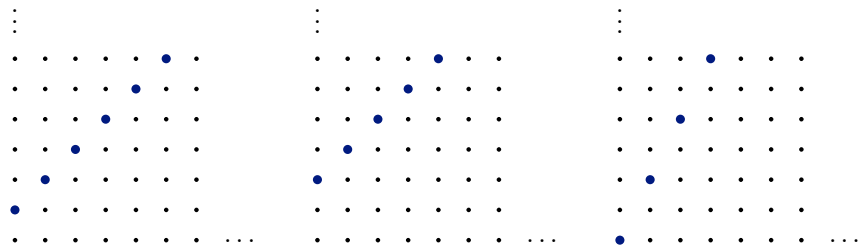


Figure 5.3: The graphs of σ, α_2 and μ_2

5.3 Maps and Subsets

If f is a map from a set A to a set B, then we have for each element a of A an image element $b \in B$. For a subset U of A, we can form the set of images of all elements of U. Thus we obtain a subset of B.

5.11 Definition. Let $f\colon A \to B$. The map $f_*\colon \mathcal{P}(A) \to \mathcal{P}(B)$ is defined by:

$$f_*(U) = \{\, f(a) \mid a \in U \,\} \qquad \text{(for all } U \subseteq A\text{)}.$$

The subset $f_*(U)$ of B is called the *image* of U under f, usually also denoted by $f(U)$. The subset $f_*(A)$ of B is called the *image* of the map f.

The image of a map is a subset of its codomain. In general it is not the codomain itself. In example 5.4 the image of f is the set $\{1, 2, 5\}$. The elements 3 and 4 of the codomain do not belong to the image. As another example, the image of $\sigma\colon \mathbb{N} \to \mathbb{N}$ is $\mathbb{N}^+$, while the codomain is $\mathbb{N}$.

In the function terminology one uses the word *range* for the image of a function.

A map $f\colon A \to B$ comes with a map $f_*\colon \mathcal{P}(A) \to \mathcal{P}(B)$. It also comes with a map in the opposite direction: from $\mathcal{P}(B)$ to $\mathcal{P}(A)$.

5.12 Definition. Let $f\colon A \to B$. The map $f^*\colon \mathcal{P}(B) \to \mathcal{P}(A)$ is defined by:

$$f^*(V) = \{\, a \in A \mid f(a) \in V \,\} \qquad \text{(for all } V \subseteq B\text{)}.$$

The subset $f^*(V)$ of A is called the *inverse image* of V under f. It is often denoted by $f^{-1}(V)$.

5.13 Example. For the map f of example 5.4 we have the following inverse images of the one element subsets of B:

$$f^*(\{1\}) = \{3\},\ f^*(\{2\}) = \{1, 2\},\ f^*(\{3\}) = f^*(\{4\}) = \emptyset \text{ and } f^*(\{5\}) = \{4\}.$$

5.4 Injective, Surjective and Bijective

5.14 Definition. A map $f\colon A \to B$ is called

- *injective* if for all $a, a' \in A$ the following holds: if $f(a) = f(a')$, then $a = a'$;
- *surjective* if for all $b \in B$ there is an $a \in A$ such that $f(a) = b$;
- *bijective* if f is both injective and surjective.

An injective map is also called an *injection*. In the same manner we speak of *surjections* and *bijections*.

Equivalent formulations:

f is injective if for every $b \in B$ the set $f^*(\{b\})$ has at most one element;

f is surjective if for all $b \in B$ the set $f^*(\{b\})$ at least one element;

f is bijective if for all $b \in B$ the set $f^*(\{b\})$ has exactly one element;

f is surjective if $f_*(A) = B$ (the image of f is the whole codomain).

5.15 Examples. The transformation $\sigma\colon \mathbb{N} \to \mathbb{N}$ is injective. That is one of Peano's axioms. The transformation $\alpha_m\colon \mathbb{N} \to \mathbb{N},\ n \mapsto n+m$ is injective as well. That is the cancellation law for the addition. The cancellation law for the multiplication tells us that $\mu_m\colon \mathbb{N} \to \mathbb{N},\ n \mapsto nm$ is injective for $m \neq 0$. Since 0 is not a successor, the map σ is not surjective. The transformation α_m is only surjective in case $m = 0$. The transformation μ_m is only surjective in case $m = 1$. In fact $\alpha_0 = \mu_1 = 1_{\mathbb{N}}$, see definition 5.18 for the definition of the transformation 1_A of a set A.

By definition, when $f\colon A \to B$ is a map, then for each $a \in A$ there is a unique element $b \in B$ such that (a, b) is in the graph of f, namely $b = f(a)$. Conversely, for an element $b \in B$ there might be more elements $a \in A$ such that $(a, b) \in \Gamma(f)$. In that case f is not injective. Or, there is not any $a \in A$ such that $(a, b) \in \Gamma(f)$, in which case f is not surjective. If a map f from A to B is bijective, then there is for each $b \in B$ a unique $a \in A$ such that $(a, b) \in \Gamma(f)$, that is $f(a) = b$. So a bijective map $f\colon A \to B$ comes with a map from B to A which maps every image element $f(a)$ back to $a \in A$.

5.16 Definition. Let $f\colon A \to B$ be bijective. The *inverse* of f is the map $f^{-1}\colon B \to A$ defined by $f^{-1}(b) = a$ if $f(a) = b$. (Note that a is the unique element of $f^*(\{b\})$.)

> Later, the -1 in the expression f^{-1} will be seen as a number, a negative number. Here it is merely part of the notation for the inverse map.

5.17 Definition. A map from a set A to itself is also called a *transformation* of A. A *permutation* is a bijective transformation.

5.18 Definition. Let A be a set. The *identity* map 1_A from A to A is determined by

$$1_A(a) = a \quad (\text{for all } a \in A).$$

1_A is also called the identity transformation of A.

> The identity transformation is a permutation. The inverse of the identity transformation is the identity transformation itself.
>
> The identity transformation maps every element to itself. That seems to be a highly uninteresting map. Nevertheless it is an important notion, comparable with the number 0 and the empty set. In case you prefer a formula for this identity transformation: the formula is x.

5.19 Definition. Let a and b be different elements of a set A. The permutation $\tau_{a,b}$ of A defined by

$$\tau_{a,b}(x) = \begin{cases} a & \text{if } x = b \\ b & \text{if } x = a \\ x & \text{otherwise.} \end{cases}$$

It is called the *transposition* of a and b.

Clearly a transposition is a permutation. The inverse of a transposition is the transposition itself: $\tau_{a,b}^{-1} = \tau_{a,b}$.

5.20 Example. The transposition of 2 and 6 in $\{1,2,3,4,5,6\}$ is the permutation

$$\tau_{2,6} = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 3 & 4 & 5 & 2 \end{smallmatrix}\right),$$

see Figure 5.4.

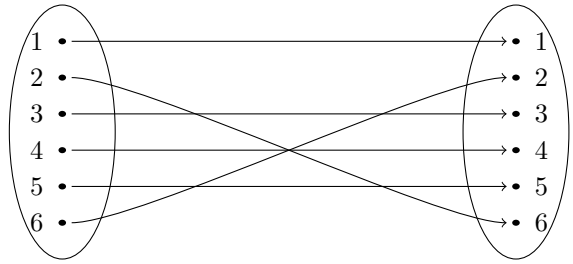


Figure 5.4: The transposition in $\{1,2,3,4,5,6\}$ of 2 and 6

Reformulation of Peano's Axioms

Using the map terminology the axioms have a compact formulation.

We have a triple $(\mathbb{N}, \sigma, 0)$ consisting of:

a) a set $\mathbb{N}$ (its elements are called *natural numbers*),
b) a transformation σ of $\mathbb{N}$ (the *successor* transformation),
c) an element 0 of $\mathbb{N}$ (the number *zero*).

The following are satisfied:

a) $0 \notin \sigma_*(\mathbb{N})$;
b) σ is injective;
c) for all $U \subseteq \mathbb{N}$: if $0 \in U$ and $\sigma_*(U) \subseteq U$, then $U = \mathbb{N}$.

5.5 The Composition of Maps

If the codomain of a map f and the domain of a map g coincide, then these maps can be composed.

5.21 Definition. Let $f\colon A \to B$ and $g\colon B \to C$ be maps. The *composition* gf of f and g is a map from A to C and is determined by

$$(gf)(a) = g(f(a)) \quad (\text{for all } a \in A).$$

The composition of f and g is also denoted by $g \circ f$. Sometimes that notation is preferred for the sake of clarity.

Notice the order in the notation of the composition: first f and then g:

$$a \mapsto f(a) \mapsto g(f(a)).$$

5.22 Example. Let $f\colon A \to B$ be the map of example 5.4. Figure 5.5 contains pictures of this map and of the map $g\colon B \to C$, where $C = \{1, 2, 3\}$ and $g(1) = g(2) = 3$, $g(3) = g(5) = 1$, $g(4) = 2$. The composition $gf\colon A \to B$ is the map with $(gf)(1) = (gf)(2) = (gf)(3) = 3$ and $(gf)(4) = 1$.

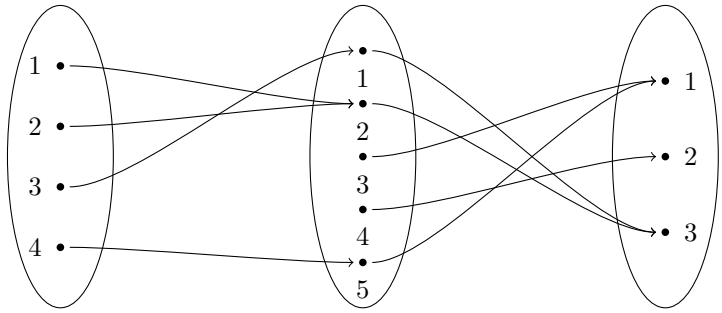


Figure 5.5: The composition of maps

5.23 Proposition. *The composition of maps is associative, i.e. let $f\colon A \to B$, $g\colon B \to C$ and $h\colon C \to D$ be maps, then*

$$h(gf) = (hg)f.$$

Proof. The maps $h(gf)$ and $(hg)f$ both have domain A and codomain D. It remains to prove that $(h(gf))(a) = ((hg)f)(a)$ for all $a \in A$. This is a direct consequence of the definition of composition: for all $a \in A$ we have

$$\begin{aligned}(h(gf))(a) &= h((gf)(a)) = h(g(f(a)))\\ ((hg)f)(a) &= (hg)(f(a)) = h(g(f(a))).\end{aligned}$$

□

Since the composition is associative there is no need for parentheses; no matter how they are placed the result is the same.

5.24 Proposition. *Let $f\colon A \to B$ be a map. Then*

$$f \circ 1_A = f = 1_B \circ f.$$

Proof. These maps have A as domain and B as codomain. Furthermore, for all $a \in A$:

$$(f \circ 1_A)(a) = f(1_A(a)) = f(a) = 1_B(f(a)) = (1_B \circ f)(a). \qquad \square$$

From the definition of the inverse of a bijection it follows that $f^{-1}f = 1_A$ and $ff^{-1} = 1_B$. Conversely these properties characterize the inverse:

5.25 Proposition. *Let $f\colon A \to B$ and $g\colon B \to A$ be maps with $gf = 1_A$ and $fg = 1_B$. Then f is bijective and $g = f^{-1}$.*

Proof. Suppose $a, a' \in A$ such that $f(a) = f(a')$. Then $(gf)(a) = (gf)(a')$, and so $a = a'$. Hence f is injective. For $b \in B$ we have $f(g(b)) = (fg)(b) = b$. So b is the image of $g(b)$ under f. It follows that f is surjective. Since f is bijective, the inverse map f^{-1} exists. For this inverse we have $f^{-1} = (gf)f^{-1} = g(ff^{-1}) = g$. $\square$

> If for example a transformation τ of a set A satisfies $\tau\tau = 1_A$, then it follows that τ is a permutation having τ itself as inverse.

The next proposition describes how injectivity and surjectivity behave under the composition of maps.

5.26 Proposition. *Let $f\colon A \to B$ and $g\colon B \to C$ be maps. Then:*

(i) *if f and g are injective, then gf is injective,*
(ii) *if gf is injective, then f is injective,*
(iii) *if f and g are surjective, then gf is surjective,*
(iv) *if gf is surjective, then g is surjective,*
(v) *if f and g are bijective, then gf is bijective and we have $(gf)^{-1} = f^{-1}g^{-1}$,*
(vi) *if gf is bijective, then f is injective and g is surjective.*

Proof.

(i) Suppose f and g are injective and let $a, a' \in A$ be such that $(gf)(a) = (gf)(a')$. Then $g(f(a)) = g(f(a'))$ and so $f(a) = f(a')$, since g is injective. It follows that $a = a'$, since f is injective.
(ii) Suppose gf is injective and let $a, a' \in A$ be such that $f(a) = f(a')$. Then $g(f(a)) = g(f(a'))$ and so $a = a'$, since gf is injective.
(iii) Suppose f and g are surjective and let $c \in C$. Then there is a $b \in B$ with $g(b) = c$, since g is surjective. Because f is surjective, there is an $a \in A$ with $f(a) = b$. Then we have $(gf)(a) = g(f(a)) = g(b) = c$. Hence gf is surjective.
(iv) Suppose gf is surjective and let $c \in C$. Because gf is surjective, there is an $a \in A$ with $(gf)(a) = c$. Then there is a $b \in B$ with $g(b) = c$, namely $b = f(a)$.
(v) The first part follows from (i) and (iii). The second part from proposition 5.25 together with $f^{-1}g^{-1}gf = f^{-1}f = 1_A$ and $gff^{-1}g^{-1} = gg^{-1} = 1_C$.
(vi) This follows from (ii) and (iv). $\square$

5.6 Numbers of Elements

Under a bijective map $f\colon A \to B$ the elements of A correspond to the elements of B. So in a sense the sets A and B have the same number of elements. However, we have not defined what the 'number of elements' actually means.

5.27 Definition. Sets A and B are called *equipotent* (or *equipollent*, or *equinumerous*, or *equivalent*) if there exists a bijection from A to B. Notation: $A \approx B$.

The notion equipotent satisfies properties one can expect:

5.28 Proposition. *Let A, B and C be sets. Then:*

(i) $A \approx A$,
(ii) *if* $A \approx B$, *then* $B \approx A$,
(iii) *if* $A \approx B$ *and* $B \approx C$, *then* $A \approx C$.

Proof.

(i) 1_A is a bijection from A to A.
(ii) If f is a bijection from A to B, then f^{-1} is a bijection from B to A.
(iii) If f is a bijection from A to B and g a bijection from B to C, then gf is a bijection from A to C. □

In order to indicate the number of elements of a set, the set can be compared with a standard set. For each natural number n we have a standard set $\underline{n}$.

5.29 Notations. Let $n \in \mathbb{N}$. By $\underline{n}$ we denote the set of natural numbers from 1 up to n:

$$\underline{n} = \{\, k \in \mathbb{N} \mid 1 \leq k \leq n \,\} = \{1, 2, \ldots, n\}.$$

We also use the following notation:

$$\mathbb{N}_n = \{\, k \in \mathbb{N} \mid k < n \,\} = \{0, 1, \ldots, n-1\}.$$

Thus we have $\underline{0} = \emptyset$ and for all $n \in \mathbb{N}$: $\underline{n+1} = \underline{n} \cup \{n+1\}$. This determines $\underline{n}$ inductively, just as the $\mathbb{N}_n$ are determined by $\mathbb{N}_0 = \emptyset$ and $\mathbb{N}_{n+1} = \mathbb{N}_n \cup \{n\}$. Clearly $\mathbb{N}_n \approx \underline{n}$.

The idea is that $\underline{n}$ is the standard set having n elements. Then a set A has n elements if $A \approx \underline{n}$. Both $A \approx \underline{m}$ and $A \approx \underline{n}$ cannot hold for different m and n. But why not? We will give a proof.

5.30 Proposition. *For all natural numbers m and n: if there exists an injective map $f\colon \underline{m} \to \underline{n}$, then $m \leq n$.*

Proof. We prove by mathematical induction that the following holds for all natural numbers n.

> $P(n)$: for all $m \in \mathbb{N}$: if there exists an injection $f\colon \underline{m} \to \underline{n}$, then $m \leq n$.

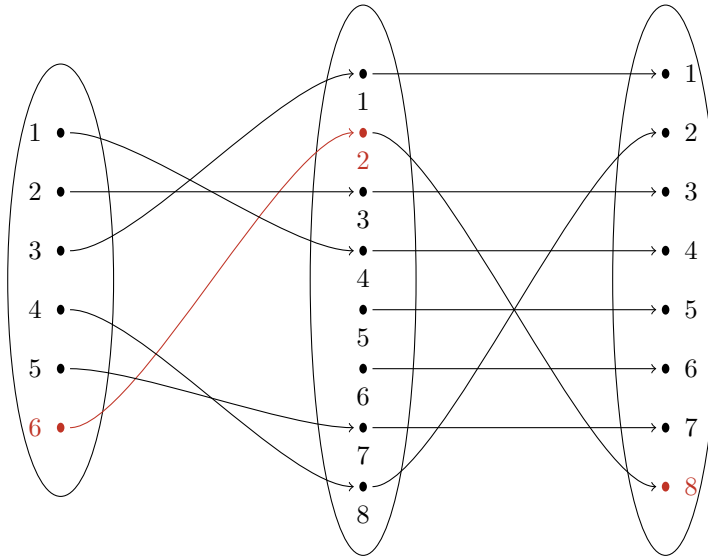


Figure 5.6: From an injection $\underline{6} \to \underline{8}$ to an injection $\underline{5} \to \underline{7}$

If there is an injection $\underline{m} \to \emptyset$, then $m = 0$. So $P(0)$ holds.

Suppose n is a natural number such that $P(n)$. We will prove that then also $P(n+1)$. Let $f\colon \underline{m} \to \underline{n+1}$ be an injective map. Because $0 \leq n+1$ we may assume that $m \neq 0$. Let $g\colon \underline{n+1} \to \underline{n+1}$ be defined by

$$g(k) = \begin{cases} f(m) & \text{if } k = n+1 \\ n+1 & \text{if } k = f(m) \\ k & \text{otherwise.} \end{cases}$$

Otherwise put: if $f(m) = n+1$, then $g = 1_{\underline{n+1}}$, and if $f(m) \neq n+1$, then $g = \tau_{f(m),n+1}$. Since g is injective, gf is injective as well. We have $(gf)(m) = g(f(m)) = n+1$ and, since gf is injective, gf maps the subset $\underline{m-1}$ into $\underline{n}$, see also Figure 5.6. Thus by restriction we have an injective map from $\underline{m-1}$ to $\underline{n}$. From $P(n)$ it follows that $m-1 \leq n$, that is $m \leq n+1$.

Hence $P(n)$ holds for all natural numbers n. □

5.31 Corollary. *Let m and n be natural numbers and let A be a set such that $A \approx \underline{m}$ and $A \approx \underline{n}$. Then $m = n$.*

Proof. From $A \approx \underline{m}$ and $A \approx \underline{n}$ it follows that $\underline{m} \approx \underline{n}$. So there is a bijection $f\colon \underline{m} \to \underline{n}$. Both f and f^{-1} are injective. Hence $m \leq n$ and $n \leq m$, that is $m = n$. □

This corollary justifies the following definition.

5.32 Definition. A set A *has* n *elements* if $A \approx \underline{n}$. Notation: $\#(A) = n$. We call n the *number* of elements of A. If such an n exists then the set A is said to be *finite*. A bijective map $\underline{n} \to A$ is said to *assign* numbers to the elements of A.

> Counting elements of a set is a very natural thing to do and many properties of counting we see as obvious. We used this already in some examples and some exercises. This section however is about its mathematical foundation.

In definition 5.32 it is defined what it means for a set to be finite. A consequence of giving a definition for a notion that is intuitively clear, is that obvious properties require a proof, as is the case for the following proposition.

5.33 Proposition. *Let B be a subset of a finite set A. Then B is finite and $\#(B) \leq \#(A)$.*

Proof. First we prove that subsets of A are finite. We do so by induction on $\#(A)$. For $\#(A) = 0$ it is obvious.

> Suppose that for an $n \in \mathbb{N}$ it is true that subsets of a set with n elements are finite. Let A be a set with $n+1$ elements. Let $f\colon \underline{n+1} \to A$ assign numbers to the elements of A. It determines an assignment $\underline{n} \to A \setminus \{f(n+1)\}$ of numbers to the elements of $A \setminus \{f(n+1)\}$ and so $\#(A \setminus \{f(n+1)\}) = n$. Let B be a subset of A. The subset $B \setminus \{f(n+1)\}$ of $A \setminus \{f(n+1)\}$ is finite. There are two possibilities: $f(n+1) \in B$ and $f(n+1) \notin B$. In both cases it follows easily that B is finite.

By mathematical induction it follows that subsets of finite sets are finite. Now let B be a subset of a set A with $\#(A) = n$. Then B is finite, say $\#(B) = m$. There are bijections $f\colon A \to \underline{n}$ and $g\colon \underline{m} \to B$. The map $\underline{m} \to \underline{n},\ x \mapsto f(g(x))$ is injective and so it follows from proposition 5.30 that $m \leq n$. □

Yet another consequence of proposition 5.30:

5.34 Corollary. *For all natural numbers m and n: if there exists a surjective map $f\colon \underline{m} \to \underline{n}$, then $m \geq n$.*

Proof. Suppose $f\colon \underline{m} \to \underline{n}$ is surjective. Choose for every $b \in \underline{n}$ an $a_b \in \underline{m}$ such that $f(a_b) = b$. Thus we have a map $g\colon \underline{n} \to \underline{m},\ b \mapsto a_b$. Then $fg = 1_{\underline{n}}$ and so the map $\underline{n} \to \underline{m},\ b \mapsto a_b$ is injective. From proposition 5.30 it follows that $n \leq m$. □

5.7 Some Counting Principles

Some counting principles that are widely used will be derived in this section.

Johann Peter Gustav Lejeune Dirichlet (Düren 1805 – Göttingen 1859)

One of Dirichlet's interests was number theory. He gave a proof for Fermat's Last Theorem for the exponents 5 and 14, see subsection 10.4.2 for Fermat's Last Theorem. A well-known theorem of Dirichlet states that in an arithmetic progression $a, a+b, a+2b, a+3b, \ldots$ of natural numbers there is a prime number if a and b have only 1 as a common factor. In the proof complex numbers are used.

5.35 Theorem (Dirichlet's principle). *Let A and B be finite sets such that $\#(A) > \#(B)$. Let f be a map from A to B. Then f is not injective. (To put it differently: there is an element $b \in B$ such that $f^*(\{b\})$ has more than one element.)*

Proof. Put $\#(A) = m$ and $\#(B) = n$. There are bijections $g\colon \underline{m} \to A$ and $h\colon B \to \underline{n}$.

> Suppose f is injective. Then $hfg\colon \underline{m} \to \underline{n}$ is also injective. However $m > n$. This contradicts proposition 5.30.

Hence f is not injective. □

> Dirichlet's principle is also known as the 'pigeonhole principle' and as the 'Schubfachprinzip', the German name Dirichlet gave to the principle: if you put objects in drawers ('Schubfachen' in German) and there are more objects than drawers, then there will be a drawer containing more than one object.

We also have:

5.36 Theorem. *Let A and B be finite sets such that $\#(A) < \#(B)$. Let f be a map from A to B. Then f is not surjective. (Alternatively: there is an element $b \in B$ such that $f^*(\{b\})$ is empty.)*

Proof. Put $\#(A) = m$ and $\#(B) = n$. There are bijections $g\colon \underline{m} \to A$ and $h\colon B \to \underline{n}$.

> Suppose f is surjective. Then $hfg\colon \underline{m} \to \underline{n}$ is also surjective. However $m < n$. This contradicts Corollary 5.34.

Hence f is not surjective. □

5.37 Theorem. *Let $f\colon A \to B$ be a map where A and B are finite sets such that $\#(A) = \#(B)$. Then:*

$$f \text{ is injective} \iff f \text{ is surjective.}$$

Proof.

$\Rightarrow$: Suppose f is not surjective. Then there is an element $b \in B$ such that $b \notin f_*(A)$. The map $A \to B \setminus \{b\}$, $a \mapsto f(a)$ is injective as well, while $\#(B \setminus \{b\}) = \#(B) - 1 < \#(A)$. In contradiction with proposition 5.35.

$\Leftarrow$: Suppose f is not injective. Then there are $a_1, a_2 \in A$ such that $a_1 \neq a_2$ and $f(a_1) = f(a_2)$. Then $A \setminus \{a_2\} \to B$, $a \mapsto f(a)$ is surjective as well, while $\#(A \setminus \{a_2\}) = m - 1 < \#(B)$. In contradiction with proposition 5.36. □

In particular a transformation $f\colon A \to A$ of a finite set A is injective if and only if it is surjective. For an infinite (= not finite) set the situation is completely different. The map $\sigma\colon \mathbb{N} \to \mathbb{N}$ for example is injective, but not surjective: the injectivity is an axiom and another axiom states that 0 is not in the image. Since $\sigma_*(\mathbb{N}) = \mathbb{N}^+$ we have a bijection from $\mathbb{N}$ to $\mathbb{N}^+$. So $\mathbb{N} \setminus \{0\} \approx \mathbb{N}$. After removing 0 from $\mathbb{N}$ the same 'number' of elements remain! In later chapters we will consider more infinite sets.

5.38 Definition. A set A is called *countable* if $A \approx \mathbb{N}$.

Countable sets are infinite. For a countable set A there is a bijection $\mathbb{N} \to A$, $n \mapsto a_n$. Thus a sequence containing all elements of A can be formed, which is without repetition. There are other (larger) infinite sets. They are called *uncountable*. In chapter 17 we will have a closer look at these.

5.8 Operations on Numbers and Sets

Natural numbers are invented for counting, for indicating the number of elements in a finite set. Operations on natural numbers correspond to operations on sets: addition corresponds to the union, multiplication to the Cartesian product.

5.39 Lemma. *Let m and n be natural numbers. Then $\underline{n} \approx m + \underline{n}$. (The notation $m + \underline{n}$ stands for the set $\{\, m + k \mid k \in \underline{n} \,\}$.)*

Proof. The map $\underline{n} \to m+\underline{n}$, $k \mapsto m+k$ has an inverse: $m+\underline{n} \to \underline{n}$, $l \mapsto l-m$. □

The following proposition, which connects the addition of numbers to the union of sets, is obviously true. It is however fundamental for many counting principles.

5.40 Proposition. *Let A and B be disjoint finite sets. Then $A \cup B$ is also finite and $\#(A \cup B) = \#(A) + \#(B)$.*

Proof. If $\#(A) = m$ and $\#(B) = n$, then there are bijections $f\colon A \to \underline{m}$ and $g\colon B \to m + \underline{n}$. The map $h\colon A \cup B \to \underline{m+n}$ defined by

$$h(x) = \begin{cases} f(x) & \text{if } x \in A, \\ g(x) & \text{if } x \in B, \end{cases}$$

is bijective. Hence $\#(A \cup B) = m + n = \#(A) + \#(B)$. □

The multiplication of natural numbers corresponds to the Cartesian product of sets:

5.41 Proposition. *Let A and B be finite sets. Then $A \times B$ is also finite and $\#(A \times B) = \#(A) \cdot \#(B)$.*

Proof. We prove this with induction on $\#(B)$. For $\#(B) = 0$, i.e. for B empty, $A \times B$ is also empty.

Assume the proposition is true for sets B with $\#(B) = n$. Then we must prove that it is also true if $\#(B) = n+1$. So suppose $\#(B) = n+1$. Choose an element $b \in B$. We have

$$A \times B = (A \times (B \setminus \{b\})) \cup (A \times \{b\}).$$

From proposition 5.40 it follows that

$$\#(A\times B) = \#(A\times(B\setminus\{b\}))+\#(A\times\{b\}) = \#(A)\cdot n+\#(A) = \#(A)\cdot(n+1).$$

□

5.42 Notation. Let A and B be sets. The set of all maps from A to B we denote by B^A. Thus

$$B^A = \{\, f \mid f\colon A \to B \,\}.$$

Exponentiation of natural numbers corresponds to this set of maps:

5.43 Proposition. *Let A and B be finite sets. Then B^A is also a finite set and $\#(B^A) = \#(B)^{\#(A)}$.*

Proof. We prove this by induction to $\#(A)$. For $\#(A) = 0$ the set B^A has only one element: $\#(B^\emptyset) = 1$.

Assume the proposition is true if $\#(A) = n$. Then we aim to prove that it also holds for $\#(A) = n + 1$. So suppose $\#(A) = n + 1$. Choose an $a \in A$. The map

$$F\colon B^A \to B^{A\setminus\{a\}} \times B,\ f \mapsto (f', f(a)),$$

where f' is the restriction of f to $A \setminus \{a\}$, is a bijection and so it follows from proposition 5.41 that

$$\#(B^A) = \#(B^{A\setminus\{a\}} \times B) = \#(B)^n \cdot \#(B) = \#(B)^{n+1}.$$ □

In the proofs of the propositions 5.41 and 5.43 no explicit bijections were given. In chapter 8 we will look at this again. For proposition 5.41 see also the exercises 8 and 9 of this chapter.

5.9 Number of Subsets

We have already seen that the number of subsets of a set A of n elements, i.e. the number of elements of $\mathcal{P}(A)$, equals 2^n. According to the previous section this also is the number of maps from A to $\{0,1\}$. We will show that the sets $\mathcal{P}(A)$ and $\{0,1\}^A$ are equipotent by constructing a bijection between these sets, whether A is finite or not.

5.44 Definition. Let A be a set and U a subset of A. The *characteristic function* of U on A is the function

$$\chi_U \colon A \to \{0,1\},\ x \mapsto \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \notin U. \end{cases}$$

5.45 Definition. Let A be a set and $f \colon A \to \{0,1\}$. The *support* of f is the subset

$$D(f) = \{\, x \in A \mid f(x) = 1 \,\}$$

of A.

5.46 Proposition. *Let A be a set. Then $\mathcal{P}(A) \approx \{0,1\}^A$.*

Proof. We will prove that the maps $U \mapsto \chi_U$ and $f \mapsto D(f)$ are each others inverses: $D(\chi_U) = U$ for all $U \subseteq A$ and $\chi_{D(f)} = f$ for all $f \colon A \to \{0,1\}$.

Let U be a subset of A. Then for all $x \in A$:

$$x \in D(\chi_U) \iff \chi_U(x) = 1 \iff x \in U.$$

Hence $D(\chi_U) = U$.

Let $f \colon A \to \{0,1\}$. Then for all $x \in A$:

$$\chi_{D(f)}(x) = 1 \iff x \in D(f) \iff f(x) = 1.$$

Hence $\chi_{D(f)} = f$. □

5.47 Example. Let $A = \{0,1,2,3,4,5,6,7,8,9\}$. The subset $U = \{3,4,6,8,9\}$ of A corresponds to the map $f = \left(\begin{smallmatrix} 0&1&2&3&4&5&6&7&8&9 \\ 0&0&0&1&1&0&1&0&1&1 \end{smallmatrix}\right)$. We have $f = \chi_U$ and $U = D(f)$. See also exercise 6 of chapter 1.

5.48 Corollary. *Let A be a finite set. Then $\#(\mathcal{P}(A)) = 2^{\#(A)}$.*

Proof. $\#(\mathcal{P}(A)) = \#(\{0,1\}^A) = \#(\{0,1\})^{\#(A)} = 2^{\#(A)}$. □

Characteristic functions of unions and intersections of subsets are determined by the characteristic functions of these subsets.

5.49 Proposition. *Let U and V be subsets of a set A and χ_U and χ_V the characteristic functions of U and V on A. Then:*

(i) $a \mapsto \chi_U(a)\chi_V(a)$ *is the characteristic function of $U \cap V$ on A.*
(ii) $a \mapsto \chi_U(a) + \chi_V(a)$ *is the characteristic function of $U \cup V$ on A if $U \cap V = \emptyset$.* □

Exercises

1. Show that $\{\{a\},\{a,b\}\} = \{\{c\},\{c,d\}\} \iff a = c$ and $b = d$.
2. Let A be a set. How many maps are there from $\emptyset$ to A? And how many from A to $\emptyset$?
3. Let $f\colon A \to B$ and $g\colon B \to C$.
   (i) Prove that $(gf)_* = g_* f_*$.
   (ii) Prove that $(gf)^* = f^* g^*$.
4. Let $f\colon A \to B$ be surjective.
   (i) Show that $f_*\colon \mathcal{P}(A) \to \mathcal{P}(B)$ is surjective.
   (ii) Show that $f^*\colon \mathcal{P}(B) \to \mathcal{P}(A)$ is injective.
5. Let $f\colon A \to B$ be injective.
   (i) Show that $f_*\colon \mathcal{P}(A) \to \mathcal{P}(B)$ is injective.
   (ii) Show that $f^*\colon \mathcal{P}(B) \to \mathcal{P}(A)$ is surjective.
6. Let f be a map from A to B. Show that $A \approx \Gamma(f)$.
7. Let $G = (V, E)$ be a graph and $v \in V$, that is v is a vertex of the graph G. The *degree* of v is the number of $e \in E$ with $v \in e$. We denote the degree of v as $\deg(v)$, so
$$\deg(v) = \#\{\, e \in E \mid v \in e \,\}.$$
Show that, if $\#(V) \geq 2$, there are two vertices of G with the same degree.
8. Let m and n be natural numbers.
   (i) Prove that for all $x \in \mathbb{N}_m$ and all $y \in \mathbb{N}_n$ we have $xn + y \in \mathbb{N}_{mn}$. (See notation 5.29 for the notation $\mathbb{N}_n$.)
   (ii) Prove that the map $f\colon \mathbb{N}_m \times \mathbb{N}_n \to \mathbb{N}_{mn}$, $(x, y) \mapsto xn + y$ is injective.
   (iii) Prove that the map f is bijective.

9. Let A and B be finite sets with $\#(A) = m$ and $\#(B) = n$. Then there are bijections $g\colon A \to \underline{m}$ and $h\colon B \to \underline{n}$. Consider the map $F\colon A \times B \to \underline{mn}$ which is the following composition:

$$A \times B \xrightarrow{s} \underline{m} \times \underline{n} \xrightarrow{t} \mathbb{N}_m \times \mathbb{N}_n \xrightarrow{f} \mathbb{N}_{mn} \xrightarrow{u} \underline{mn},$$

where f is the bijection from exercise 8 and the maps s, t and u are determined by $s(a,b) = (g(a), h(b))$, $t(x,y) = (x-1, y-1)$ and $u(z) = z+1$.

   (i) Show that F is bijective.

   (ii) Show that $F(a,b) = (g(a)-1)n + h(b)$.

   In this manner we have not only shown that $\#(A \times B) = mn$, but we also constructed a bijection $F\colon A \times B \to \underline{mn}$ out of given bijections $g\colon A \to \underline{m}$ and $h\colon B \to \underline{n}$.

10. Let A be an infinite set. Show that there is an injective map $f\colon \mathbb{N} \to A$. (Construct such an f step by step: show that an injective $f_n\colon \mathbb{N}_n \to A$ can be extended to an injective $f_{n+1}\colon \mathbb{N}_{n+1} \to A$.)

11. Let A be an infinite set and let a be an element of A. Prove that $A \approx A \setminus \{a\}$. (Hint: use exercise 10.)

12. Let A a countable set and B an infinite subset of A. Prove that B is countable.

13. Let A be a countable set and $f\colon A \to B$ a surjective map with B infinite. Prove that B is countable.

14. Let A and B be countable sets. Prove that $A \times B$ is countable.

15. Let U and V be subsets of a set A. These subsets correspond to characteristic functions χ_U and χ_V on A.

   (i) Show that the function

$$A \to \{0,1\}, \quad a \mapsto \chi_U(a) + \chi_V(a) - 2\chi_U(a)\chi_V(a)$$

   is the characteristic function of $U \div V$. See exercise 5 of chapter 2.

   (ii) How is the characteristic function of $U \cup V$ determined by χ_U and χ_V?

16. Let $A_0, A_1, A_2, \ldots$ be countable sets with bijections $f_0\colon \mathbb{N} \to A_0$, $f_1\colon \mathbb{N} \to A_1$, $f_2\colon \mathbb{N} \to A_2$, $\ldots$

   (i) Show that the map

$$F\colon \mathbb{N} \times \mathbb{N} \to \bigcup_{n=0}^{\infty} A_n, \quad (n,m) \mapsto f_n(m)$$

   is surjective.

   (ii) So the set $\bigcup_{n=0}^{\infty} A_n$ is countable. Why?

   (iii) Show that the map F from item (i) is bijective if and only if $A_i \cap A_j = \emptyset$ for $i \neq j$.

6 Iteration

For transformations domain and codomain coincide. So transformations of the same set can always be composed. In particular a transformation can be composed with itself, and again the result can be composed with that transformation, etc. This is the so-called iteration of the transformation.

6.1 Transformations

By definition 5.17 a transformation is a map with the same set as domain and codomain.

6.1 Example. Let $A = \underline{10}$. A transformation f of A is given by the image elements $f(a)$ of all elements a of A. For example, $f(1) = 4$, $f(2) = 10$, $f(3) = 4$, $f(4) = 10$, $f(5) = 5$, $f(6) = 9$, $f(7) = 1$, $f(8) = 2$, $f(9) = 10$, $f(10) = 3$. Since domain and codomain coincide, one can make a picture of a transformation by indicating with arrows in a picture of the set A how elements are mapped. In Figure 6.1 there is a picture of this transformation.

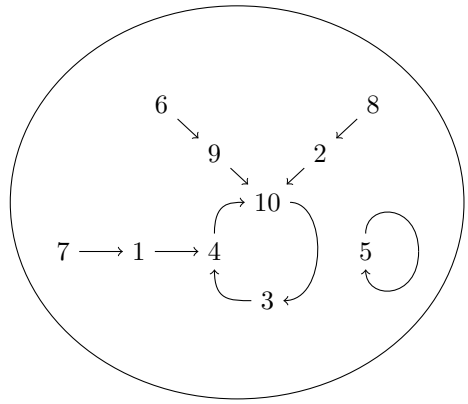


Figure 6.1: Picture of a transformation

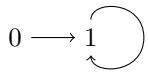
Once more Peano's Axioms

Peano's axioms concern a triple $(\mathbb{N}, \sigma, 0)$, where $\mathbb{N}$ is a set, σ a transformation of $\mathbb{N}$ and 0 an element of $\mathbb{N}$.

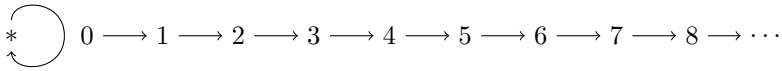
We will give three triples $(A, \sigma, 0)$ of sets A with a transformation σ and an element 0 such that in each of the three cases exactly two of the three axioms are satisfied.

a) A transformation not satisfying the first axiom, but satisfying the other two:

b) Not satisfying the second axiom, but satisfying the other two:



c) Not satisfying the third axiom, but satisfying the other two:



From this it follows that each of the three axioms is not a consequence of the other two.

6.2 Sequences and Tuples

A sequence $a_0, a_1, a_2, \ldots$ in a set A can be seen as a map. In fact one can define sequences as maps.

6.2 Definition. An *infinite sequence* in a set A is a map from $\mathbb{N}$ to A. Usually, infinite sequences are just called sequences.

6.3 Notation. If $a\colon \mathbb{N} \to A$ is a map, then every $n \in \mathbb{N}$ has an image element $a(n) \in A$. Instead of $a(n)$ one often writes a_n. The sequence is usually denoted by $a_0, a_1, a_2, \ldots$ and also by (a_n), where it is understood that the n 'varies' over the natural numbers. The image of n, that is a_n, is called the *n-th term* of the sequence.

There is no end to a sequence: there is no last term. Also sequences that do have an end can be considered:

6.4 Definition. Let $n \in \mathbb{N}$. A *sequence* of *length* n, or an *n-tuple*, in a set A is a map from $\mathbb{N}_n$ to A. It is also called a *finite sequence* without further reference to its length.

6.5 Notation. An n-tuple $a\colon \mathbb{N}_n \to A$ is often denoted by $a_0, a_1, \ldots, a_{n-1}$.

6.6 Notations. Let A be a set. For later use we introduce the following notations:

$\mathcal{R}(A)$: the set of infinite sequences in A $(= A^{\mathbb{N}})$,
$\mathcal{F}(A)$: the set of finite sequences in A,
$\mathcal{F}_n(A)$: the set of finite sequences of length n in A $(= A^{\mathbb{N}_n})$.

6.3 Recursive Definitions

Let A be a set. A sequence (a_n) in A can be given by a direct definition of the terms a_n. For example: $a_n = n^2$. In this example the sequence (a_n) is the sequence of the squares. A sequence (a_n) can also be determined in a *recursive* way: then a_n is defined in terms of $(a_0, \ldots, a_{n-1})$.

A special case is the so-called *simple recursion*. Then every term in the sequence is determined by the preceding term, except for the first term. In this way a sequence (a_n) in a set A is given by its first term a_0 and a transformation f of A:

$$\begin{cases} a_0 = a, \\ a_{n+1} = f(a_n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

6.7 Examples. The definitions of addition, multiplication and exponentiation of natural numbers in section 4.4 are examples of simple recursion.

Addition. For each $m \in \mathbb{N}$ a sequence $m+0, m+1, m+2, \ldots$ in $\mathbb{N}$ is defined: the first term is m and the transformation is σ, the successor map.

Multiplication. For each $m \in \mathbb{N}$ a sequence $m \cdot 0$, $m \cdot 1$, $m \cdot 2$, $\ldots$ in $\mathbb{N}$ is defined: the first term is 0 and the transformation is α_m, adding m.

Exponentiation. For each $m \in \mathbb{N}$ a sequence $m^0, m^1, m^2, \ldots$ in $\mathbb{N}$ is defined: The first term is 1 and the transformation is μ_m, multiplication by m.

6.8 Example. On page 35 we saw that the Tower of Hanoi is solvable for any number if discs. We saw how from a solution for n discs a solution for $n+1$ discs can be constructed. In the solution thus obtained the largest disc is moved only once. It follows that this solution is one with a minimal number of moves. If we denote the number of moves in this solution for the case of n discs by a_n, we obtain a recursively defined sequence $a_0, a_1, a_2, \ldots$:

$$\begin{cases} a_0 = 0, \\ a_{n+1} = 2a_n + 1 \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

In chapter 1 it was evident from the graphs of the Tower of Hanoi that in case of n discs $2^n - 1$ moves are needed. Can we conclude that the sequence (b_n) with $b_n = 2^n - 1$ is the same sequence as (a_n)? The sequence (a_n) is defined recursively. If the sequence (b_n) satisfies the defining conditions for (a_n), these sequences must coincide. We have to show that

$$\begin{cases} b_0 = 0, \\ b_{n+1} = 2b_n + 1 \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

This is easily verified: $b_0 = 2^0 - 1 = 0$ and $2b_n + 1 = 2(2^n - 1) + 1 = 2^{n+1} - 1 = b_{n+1}$ for all $n \in \mathbb{N}$.

Later we will often use $n!$ (n *factorial*). A recursive definition is easily given:

6.9 Definition. The natural number $n!$ is for natural numbers n defined by

$$\begin{cases} 0! = 1, \\ (n+1)! = (n+1) \cdot n! \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Thus for example $5! = 5 \cdot 4! = 5 \cdot 4 \cdot 3! = 5 \cdot 4 \cdot 3 \cdot 2! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 0! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$. Less formally one can write

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1.$$

This is not a simple recursion: the n-th term is not only determined by its predecessor, but also by its number n. However, a definition by simple recursion is possible using a transformation of $\mathbb{N} \times \mathbb{N}$ instead of a transformation of $\mathbb{N}$:

$$\mathbb{N} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N}, \quad (n, x) \mapsto (n+1, (n+1)x).$$

Then there is a sequence (n, a_n) in $\mathbb{N} \times \mathbb{N}$ with

$$\begin{cases} (0, a_0) = (0, 1), \\ (n+1, a_{n+1}) = (n+1, (n+1)a_n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Thus the sequence (a_n) is the sequence $(n!)$. In some cases each next term is determined by the two preceding terms, if they exist:

6.10 Example. The sequence $f_0, f_1, f_2, \ldots$ is defined by:

$$\begin{cases} f_0 = 0, \\ f_1 = 1, \\ f_{n+2} = f_n + f_{n+1} \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

So each next term is the sum of the last two terms; if there are no two last terms, the term is given directly. Thus the sequence is formed by starting with 0 and 1 and repeatedly adding a term by taking the sum of the two last terms:

Fibonacci (Pisa(?) 1170 – Pisa(?) 1250)

Fibonacci's real name was **Leonardo Pisano**. He contributed a lot to the revival of mathematics in Europe. He is still well-known for the Fibonacci-numbers he introduced in his book *Liber abaci*. In that book he also introduced in Europe the decimal system for the notation of numbers. This notation originated in India and the Arabic world. The Fibonacci numbers were about the number of pairs of rabbits in the n-th generation when starting with 1 pair ($f_1 = 1$) and making the assumption that every pair generates a new pair in the next generation and also in the generation thereafter, but thereafter the pair will die.

n:	0	1	2	3	4	5	6	7	8	9	10	11	$\cdots$
f_n:	0	1	1	2	3	5	8	13	21	34	55	89	$\cdots$

The number f_n is called the *n-th Fibonacci-number*. The sequence can also be given by simple recursion using the transformation $(x, y) \mapsto (y, x+y)$ of $\mathbb{N} \times \mathbb{N}$ and $(0, 1)$ as first term:

$$\begin{cases} (a_0, b_0) = (0, 1), \\ (a_{n+1}, b_{n+1}) = (b_n, a_n + b_n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Then $(a_n, b_n) = (f_n, f_{n+1})$ for all $n \in \mathbb{N}$. Fibonacci-numbers pop up in many places and they have intriguing properties. There even is a journal—The Fibonacci Quarterly—devoted to these numbers and related topics.

6.4 Iteration of Transformations

If f is a transformation of a set A, then by simple recursion a sequence (a_n) in A is defined for any $a \in A$:

$$\begin{cases} a_0 = a, \\ a_{n+1} = f(a_n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

So

$$a = a_0 \overset{f}{\mapsto} a_1 \overset{f}{\mapsto} a_2 \overset{f}{\mapsto} \cdots \overset{f}{\mapsto} a_n \overset{f}{\mapsto} a_{n+1} \overset{f}{\mapsto} \cdots$$

6.11 Definition. The sequence (a_n) is called the *course* of a under f.

6.12 Example. Let f be the transformation of example 6.1. The course of 7 under f is the sequence $7, 1, 4, 10, 3, 4, 10, 3, 4, 10, 3, 4, 10, 3, 4, 10, 3, 4, \ldots$. The course of 5 is $5, 5, 5, 5, 5, 5, 5, 5, 5, 5, \ldots$.

Python

In the Python module `combinatorics.py` we put some functions which return (a part of) the course of an element under a transformation. The function `course(f, N, a)` returns the first `N` terms of the course of the element `a` under the transformation `f`.

combinatorics.py

```
def course(f, N, a):
    seq = [a]
    i = 1
    while i < N:
      a = f(a)
      seq.append(a)
      i = i + 1
    return seq
```

Below some examples are given of transformations `f` for use with the Python function `course(f,N,a)`. The first is the transformation of example 6.1. The others are from `integer.py`. The transformation `lambda x:isum(x,7)` for example is the transformation which adds `x` to the natural number `7`.

```
>>> def fun(x):
...   return {1:4, 2:10, 3:4, 4:10, 5:5, 6:9, 7:1, 8:2, 9:10, 10:3}[x
]
...
```

```
>>> fun(7)
1
>>> course(fun, 20, 7)
[7, 1, 4, 10, 3, 4, 10, 3, 4, 10, 3, 4, 10, 3, 4, 10, 3, 4, 10, 3]
>>> from integer import *
>>> course(succ, 11, 17)
[17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]
>>> course(lambda x: isum(x, 7), 11, 25)
[25, 32, 39, 46, 53, 60, 67, 74, 81, 88, 95]
>>> course(lambda x: iprod(x, 7), 6, 25)
[25, 175, 1225, 8575, 60025, 420175]
```

Thus for each $n \in \mathbb{N}$ and $a \in A$ we have an element $a_n \in A$, that is we have a transformation $a \mapsto a_n$ of A. These transformations can also easily be defined using the composition of transformations:

6.13 Definition. Let f be a transformation of A. The transformations f^n with $n \in \mathbb{N}$ of A are defined by

$$\begin{cases} f^0 = 1_A, \\ f^{n+1} = ff^n \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

The transformation f^n of A is called the *n-th iterate* of f.

> Another word for 'transformation' is '*operator*'. One also says that a transformation of A (= operator on A) *operates* on the elements of A, or that the transformation is *applied* to elements of A. Iteration of a transformation is the repeated application of the transformation.

6.14 Proposition. *Let f be a transformation of A and $a \in A$. Then the sequence $(f^n(a))$ is the course of a under the transformation f.*

Proof. The sequence $(f^n(a))$ satisfies the definition of the course:

$$\begin{aligned} f^0(a) &= 1_A(a) = a, \\ f^{n+1}(a) &= (ff^n)(a) = f(f^n(a)). \end{aligned}$$

□

6.15 Definition. A couple (A, f) consisting of a set A and a transformation f of that set is sometimes referred to as a *(discrete) dynamical system.*

> Calling such a couple (A, f) a dynamical system indicates one's interest in the course of the elements under f. In fact it is just a transformation. There also is the suggestion that one is dealing with a process in time: one starts with an a at time 0 and goes through the sequence $f^n(a)$, at time n being at $f^n(a)$.

6.16 Examples. Using iterates of transformations one can see the operations of addition, multiplication and exponentiation as repeatedly applying a transformation:

Addition is repeatedly taking the successor: $m + n = \sigma^n(m)$.

Multiplication is repeated addition: $mn = \alpha_m^n(0)$.

Exponentiation is repeated multiplication: $m^n = \mu_m^n(1)$.

6.17 Proposition (Rules for iterates). *Let f be a transformation of a set A, let m and n be natural numbers and let g be a transformation of A which commutes with f, that is $fg = gf$. Then*

(i) $f^n f^m = f^{m+n}$.
(ii) $(f^m)^n = f^{mn}$.
(iii) $(gf)^n = g^n f^n$.

Proof.

(i) By induction on n. The induction step is as follows: $f^{n+1} f^m = ff^n f^m = ff^{m+n} = f^{m+n+1}$.

(ii) By induction on n. The induction step: $(f^m)^{n+1} = f^m(f^m)^n = f^m f^{mn} = f^{mn+m} = f^{m(n+1)}$.

(iii) First prove by induction on n that f commutes with g^n. The induction step is here: $fg^{n+1} = fgg^n = gfg^n = gg^n f = g^{n+1}f$. Next to prove by induction on n that $(gf)^n = g^n f^n$. The induction step: $(gf)^{n+1} = gf(gf)^n = gfg^n f^n = gg^n ff^n = g^{n+1}f^{n+1}$. □

These rules were proved by induction, but it is also important to understand the rules intuitively. The notation f^n stands for the composition of n times the transformation f, so $f^n = \overbrace{f \dots f}^{n}$. Then the first rule is obvious:

$$f^n f^m = \overbrace{f \dots f}^{n}\overbrace{f \dots f}^{m} = \overbrace{f \dots f}^{m+n}.$$

Also the other rules can be understood this way.

Python

The function `iterate(f, n, a)` returns the image of `a` under the `n`-th iterate of `f`.

combinatorics.py

```
def iterate(f, n, a):
    i = 0
    while i != n:
        a = f(a)
        i = i + 1
    return a
```

```
>>> iterate(fun, 12, 7)
10
>>> iterate(lambda x:isum(x, 12), 45, 123)
663
```

6.5 Repeating Sequences

Sequences which occur as the course of an element under a transformation have a special shape: when two terms in the sequence happen to be equal, the sequence is repeating.

6.18 Definition. A sequence (a_n) in a set A is called *repeating* (or *eventually periodic*) if there are an $m \in \mathbb{N}$ and a $k \in \mathbb{N}^+$ such that $a_n = a_{n+k}$ for all $n \geq m$. We also say that the sequence *repeats* from the m-th term. The finite sequence $a_m, \dots, a_{m+k-1}$ we call a *period* of length k of the sequence. The finite sequence

$a_0, \ldots, a_{m-1}$ we call the *initial part* preceding the period $a_m, \ldots, a_{m+k-1}$. The sequence is called *purely repeating* (or *periodic*) if moreover $m = 0$.

If the sequence (a_n) repeats from the m-th term with a period of length k, the sequence is often notated as follows

$$a_0, a_1, a_2, \ldots, a_{m-1}, \overline{a_m, \ldots, a_{m+k-1}}.$$

The number sequence $1, 2, 3, 4, 5, 6, 7, 4, 5, 6, 7, 4, 5, 6, 7, 4, 5, 6, 7, 4, 5, 6, 7, \ldots$ has a period $4, 5, 6, 7$ of length 4 with $1, 2, 3$ as initial part:

$$1, 2, 3, \overline{4, 5, 6, 7}.$$

Or a period $5, 6, 7, 4$ of length 4 with initial part $1, 2, 3, 4$:

$$1, 2, 3, 4, \overline{5, 6, 7, 4}.$$

Or a period $5, 6, 7, 4, 5, 6, 7, 4, 5, 6, 7, 4$ of length 12 with initial part $1, 2, 3, 4$:

$$1, 2, 3, 4, \overline{5, 6, 7, 4, 5, 6, 7, 4, 5, 6, 7, 4}.$$

There are infinitely many possibilities. There exist, in the notation used in the definition, least numbers m and k. Here they are respectively 3 and 4.

6.19 Proposition. *Let f be a transformation of a set A and let the sequence (a_n) be the course of the element $a \in A$. Then, either $n \mapsto a_n$ is injective, that is all terms are different, or the sequence (a_n) repeats.*

Proof. Suppose $n \mapsto a_n$ is not injective. Then we must prove that (a_n) repeats. Since $n \mapsto a_n$ is not injective, there exist $n_1, n_2 \in \mathbb{N}$ with $n_1 < n_2$ and $a_{n_1} = a_{n_2}$. Put $k = n_2 - n_1$. Then the sequence repeats from the n_1-th term with a period of length k, because for all $l \in \mathbb{N}$:

$$a_{n_1+l} = f^{n_1+l}(a) = f^l f^{n_1}(a) = f^l(a_{n_1}) = f^l(a_{n_1+k}) = f^l f^{n_1+k}(a) = a_{n_1+l+k}. \quad \square$$

6.20 Corollary. *Let f be a transformation of a finite set A. Then for all elements of A the course under f repeats.*

Proof. The course is not injective, because A has only finitely many transformations. $\square$

6.21 The $3n + 1$ conjecture. Here is a simple transformation f of $\mathbb{N}^+$ for which the nature of the course of elements is an open problem:

$$f(n) = \begin{cases} \frac{3n+1}{2} & \text{if } n \text{ is odd,} \\ \frac{n}{2} & \text{if } n \text{ is even.} \end{cases}$$

The course of 1: $\overline{1, 2}$.
The course of 3: $3, 5, 8, 4, \overline{2, 1}$.

The course of 6: $6, 3, 5, 8, 4, \overline{2, 1}$.
The course of 7: $7, 11, 17, 26, 13, 20, 10, 5, 8, 4, \overline{2, 1}$.
The conjecture is that the course of every number repeats with period $\overline{2, 1}$. For numbers up to $10 \cdot 2^{58}$ this has been verified by computer, but still infinitely many numbers remain for which it is unknown. The conjecture is also known as the Collatz conjecture, the Ulam conjecture, the Syracuse conjecture. It probably was first formulated by Collatz in 1937.

Python

If the course repeats, for its computation you can stop as soon as a term equals a previous term. The function `repcourse(f,a)` returns for a transformation with a repeating course the smallest initial part and the smallest period.

combinatorics.py

```
def repcourse(f, a):
    pper = []
    while a not in pper:
        pper.append(a)
        a = f(a)
    i = pper.index(a)
    return [pper[:i], pper[i:]]
```

```
>>> repcourse(fun, 7)
[[7, 1], [4, 10, 3]]
>>> repcourse(fun, 3)
[[], [3, 4, 10]]
>>> repcourse(fun, 5)
[[], [5]]
```

The function `collatz(n)` is the transformation from example 6.21. Here we use the natural numbers as they occur in python: the data type `integer` together with its methods.

```
>>> def collatz(n):
...   if (n % 2) == 1: return (3 * n + 1) // 2
...   else: return n // 2
...
>>> collatz(721)
1082
>>> collatz(722)
361
>>> repcourse(collatz, 67)
[[67, 101, 152, 76, 38, 19, 29, 44, 22, 11, 17, 26, 13, 20, 10, 5, 8,
 4], [2, 1]]
```

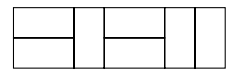
```
>>> repcourse(collatz, 2**10 - 1)
[[1023, 1535, 2303, 3455, 5183, 7775, 11663, 17495, 26243, 39365, 590
48, 29524, 14762, 7381, 11072, 5536, 2768, 1384, 692, 346, 173, 260,
130, 65, 98, 49, 74, 37, 56, 28, 14, 7, 11, 17, 26, 13, 20, 10, 5, 8,
 4], [2, 1]]
```

Exercises

1. The transformation f of the set $\underline{10}$ is given by $f(1) = 2$, $f(2) = 3$, $f(3) = 4$, $f(4) = 1$, $f(5) = 6$, $f(6) = 7$, $f(7) = 8$, $f(8) = 9$, $f(9) = 10$ and $f(10) = 5$.
   (i) Draw a picture of this transformation.
   (ii) What is the course of 8 under f? And of 2?
   (iii) Draw pictures of f^2 and f^3.
   (iv) The sequence $f^0, f^1, f^2, f^3, \ldots$ repeats. What is the length of the smallest period?

2. (i) Let f be a transformation of a finite set A. Show that the sequence $f^0, f^1, f^2, f^3, \ldots$ repeats.
   (ii) What is the length of the smallest period of this sequence if f is the transformation of example 6.1.

3. We fill rectangles of width 2 and length n ($2 \times n$-rectangles) with 2×1-rectangles. A tessellation of a 2×7-rectangle is for example:



   Let a_n be the number of ways a $2 \times n$-rectangle can be filled with 2×1-rectangles.
   (i) Determine a_1, a_2, a_3 and a_4.
   (ii) What is a_n?

4. The sequence (g_n) is defined in the same way as the sequence of Fibonacci numbers, but with other initial values:

$$\begin{cases} g_0 = 1, \\ g_1 = 0, \\ g_{n+2} = g_n + g_{n+1} \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

   (i) What is the relation between the numbers g_n and the Fibonacci numbers f_n?
   (ii) Let more generally the sequence (x_n) be defined with the initial values x and

y.

$$\begin{cases} x_0 = x, \\ x_1 = y, \\ x_{n+2} = x_n + x_{n+1} \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Describe a relation between the numbers x_n and the Fibonacci numbers f_n.

5. The sequence (s_n) of natural numbers is defined by

$$\begin{cases} s_0 = 0, \\ s_{n+1} = s_n + 2n + 1 \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Give a formula for s_n and prove its correctness.

6. Let $b \in \mathbb{N}$. The sequence (a_n) is defined by:

$$\begin{cases} a_0 = b, \\ a_{n+1} = 2a_n + 1 \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

For $b = 0$ this is the sequence of example 6.8. Determine a formula for a_n.

7. A variation on the transformation of the $3n + 1$ conjecture, see example 6.21. Consider the transformation g of $\mathbb{N}^+$ defined by

$$g(n) = \begin{cases} \dfrac{3n-1}{2} & \text{if } n \text{ is odd,} \\ \dfrac{n}{2} & \text{if } n \text{ is even.} \end{cases}$$

Can a '$3n - 1$ conjecture' be formulated?

8. The sequence $d_0, d_1, d_2, \ldots$ of natural numbers is defined by

$$\begin{cases} d_0 = 0, \\ d_{n+1} = d_n + n + 1 \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

The numbers d_n one calls, for obvious reasons, *triangular numbers*. Determine a formula for d_n.

9. Show that the map $\mathbb{N}^2 \to \mathbb{N}$, $(m, n) \mapsto \frac{(m+n)^2+3m+n}{2}$ is bijective. (Hint: compute the images for small m and n and use the triangular number d_{m+n} from exercise 8.)

10. We define the natural numbers $n?$ for $n \in \mathbb{N}$ by:

$$\begin{cases} 0? = 1, \\ (n+1)? = (2n+1) \cdot n? \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Prove that $2^n \cdot n! \cdot (n+1)? = (2n+1)!$ for all $n \in \mathbb{N}$.

11. The transformation g of $\mathbb{N}^2$ is defined by

$$g(a, b) = (2b + 1, 3a + 2) \qquad (\text{for all } a, b \in \mathbb{N}).$$

(i) Is g injective? Is g surjective?

(ii) Show that $g^{2n}(0,0) = (6^n - 1, 6^n - 1)$ for all $n \in \mathbb{N}$.

12. The transformation τ of $\mathbb{N}$ is defined by

$$\tau(n) = \begin{cases} n-1 & \text{if } n \geq 1, \\ 0 & \text{if } n = 0. \end{cases}$$

(i) Show that $\tau\sigma = 1_{\mathbb{N}}$ and $\sigma\tau \neq 1_{\mathbb{N}}$. (Here $\sigma(n)$ is the successor of n.)

(ii) Give a transformation τ' of $\mathbb{N}$ with $\tau'\sigma = 1_{\mathbb{N}}$ and $\tau' \neq \tau$.

(iii) Prove that $\tau^k\sigma^k = 1_{\mathbb{N}}$ for all $k \in \mathbb{N}$.

(iv) Prove that for every $n \in \mathbb{N}$ there is a $k \in \mathbb{N}$ such that $\sigma^k\tau^k(n) \neq n$.

13. Let f_n be the n-th Fibonacci number. Prove that for all $n \in \mathbb{N}$

$$\sum_{k=0}^{n} {f_k}^2 = f_n f_{n+1}.$$

14. Let $f_0, f_1, \ldots$ be the sequence of Fibonacci numbers.

(i) Prove that for all $n, k \in \mathbb{N}$ with $n \geq k$

$$f_{n+1} = f_{k+1}f_{n+1-k} + f_k f_{n-k}.$$

(ii) Prove that for all $m \in \mathbb{N}$ we have $f_{2m+1} = f_{m+1}^2 + f_m^2$.

15. The transformation $f\colon \mathbb{N}^+ \to \mathbb{N}^+$ is defined by

$$f = \begin{cases} n+2 & \text{if } n \text{ is not divisible by 3,} \\ \dfrac{n}{3} & \text{if } n \text{ is divisable by 3.} \end{cases}$$

(i) Prove that for every $n \in \mathbb{N}^+$ there is a $k \in \mathbb{N}$ such that $f^k(n) \leq 2$.

(ii) Is there a $k \in \mathbb{N}$ such that $f^k(n) \leq 2$ for every $n \in \mathbb{N}^+$?

7 The Integers

In this chapter we will extend $\mathbb{N}$ (the natural numbers) with negative numbers to $\mathbb{Z}$ (the integers). As far as it concerns only the *set* of integers this is not hard to do: we might simply add new elements -1, -2, -3, ... to $\mathbb{N}$. But when we want to extend addition and multiplication of natural numbers to the set of all integers, this procedure leads to distinguishing many cases, in particular when deriving the rules of arithmetic. That is why we proceed differently.

What do we want to achieve? In $\mathbb{N}$ equations

$$m + x = n \tag{7.1}$$

do not have a solution in general. There is a unique solution $n - m$ if $m \leq n$ and there is no solution if $m > n$. We want to extend $\mathbb{N}$ to a set $\mathbb{Z}$ and moreover extend the addition in $\mathbb{N}$ to an addition in $\mathbb{Z}$ in such a way that:

a) equations like (7.1) have a unique solution for any $m, n \in \mathbb{Z}$;
b) the rules for addition of natural numbers are extended to rules for addition of integers;
c) every 'new' number is needed to fulfill the two conditions above.

Having achieved this, we proceed to extend the multiplication in $\mathbb{N}$ and the ordering of $\mathbb{N}$ to $\mathbb{Z}$. Equation (7.1) must have a solution for any $m, n \in \mathbb{N}$. Such a solution is an integer. Thus every ordered pair (n, m) determines an integer. We will see integers as differences of natural numbers. Because we would like to do arithmetic with the integers as we do with the natural numbers, we are forced to see differences that represent integers as being equal: if $n_1 - m_1 = n_2 - m_2$, then $n_1 + m_2 = m_1 + n_2$, and what this means we know, since it refers only to arithmetic in $\mathbb{N}$. We will introduce $\mathbb{Z}$ by grouping all ordered pairs (n, m) into classes and those classes will be integers by definition. In order to follow this program we first set up the basic mathematical machinery for such constructions. Here it might appear to be somewhat overdone, but this machinery is used in mathematics over and over again. It is worthwhile to become familiar with it at an early stage.

7.1 Partitions

A map $f\colon A \to B$ determines the image $f_*(A)$ of f: the subset of B consisting of all images under f of elements of A. By replacing the codomain B of f by the

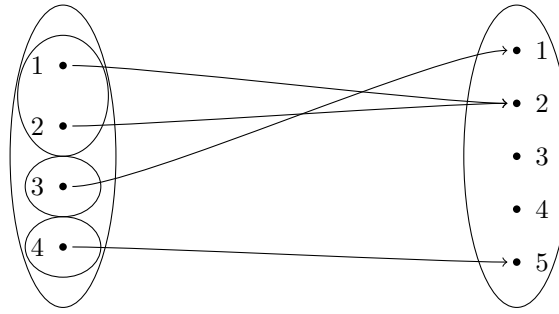


Figure 7.1: Picture of an f-partition

image $f_*(A)$ we obtain a surjective map $A \to f_*(A)$, $a \mapsto f(a)$. Thus we forced the map to become surjective by adjusting the codomain. The resulting map is bijective if and only if f is injective.

By adjusting the domain A of a map $f\colon A \to B$ we can force the map to become injective in such a way that the resulting map is bijective if and only if f is surjective. In a way we do this by seeing elements of A having the same image under f as being equal.

7.1 Definition. Let $f\colon A \to B$ be a map. For every $a \in A$ there is the subset of A of all elements which map under f to the same element as a:

$$[a]_f = \{\, x \in A \mid f(x) = f(a) \,\}$$

This subset of A is called the *f-class* of a. Let A_f be the set of all these f-classes:

$$A_f = \{\, [a]_f \mid a \in A \,\}.$$

This set of classes is called the *f-partition* of A.

7.2 Example. The map $f = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 1 & 5 \end{smallmatrix}\right)\colon \underline{4} \to \underline{5}$, see example 5.4, gives the following f-partition of $\underline{4}$:

$$\underline{4}_f = \{\{1, 2\}, \{3\}, \{4\}\},$$

See Figure 7.1.

Clearly the f-partition of A has the following properties:

a) f-classes are not empty.
b) For all $a \in A$ there is a unique f-class containing a.

As we associated to a map $f\colon A \to B$ a surjective map $A \to f_*(A)$, $a \mapsto f(a)$, we now also have an injective map $A_f \to B$, $[a]_f \mapsto f(a)$. Thus f is a composition of three maps:

$$A \to A_f \to f_*(A) \to B.$$

The first is surjective, the second bijective and the third is injective. Is f surjective, then $A_f \to B$ is a bijection.

7.3 Example. The map of example 7.2 induces a bijection

$$\{\{1,2\},\{3\},\{4\}\} \to \{1,2,5\},$$

where $\{1,2\} \mapsto 2$, $\{3\} \mapsto 1$ and $\{4\} \mapsto 5$.

An f-partition of A is a subdivision of the elements of A into classes of elements having equal images under f. For describing a subdivision into classes a map f need not to be given.

7.4 Definition. Let A be a set. A set Φ of subsets of A is called a *partition* of A if:

a) $\emptyset \notin \Phi$,
b) for all $a \in A$ there is a unique $U \in \Phi$ such that $a \in U$.

The sets $U \in \Phi$ we call *classes* of the partition. If $a \in U$, where $U \in \Phi$, then we call a a *representative* of U. We also say that U is the *class* of a; notation $U = [a]_\Phi$. A subset R of A having precisely one element in common with each of the classes $U \in \Phi$, we call a *system of representatives* of Φ.

7.5 Example. In example 7.3 we have the partition $\{\{1,2\},\{3\},\{4\}\}$ of $\{1,2,3,4\}$ consisting of 3 classes. A system of representatives is $\{1,3,4\}$. Also $\{2,3,4\}$ is one.

> In the second requirement for a partition the word unique occurs. This can be split into two requirements:
>
> a) Every element of A is in at least one of the classes of Φ, that is the union of all classes is the set A: $\bigcup_{U\in\Phi} U = A$.
> b) Every element of A is in at most one of the classes of Φ, that is two different classes are disjoint: $U \cap V = \emptyset$ for all $U, V \in \Phi$ with $U \neq V$.

Since for every $a \in A$ there is a unique $U \in \Phi$ such that $a \in U$, a map $A \to \Phi$ is defined by $a \mapsto U$. Because $\emptyset \notin \Phi$ this map is surjective. Thus a partition Φ of A determines a surjective map $A \to \Phi$, $a \mapsto [a]_\Phi$. Notice the analogy with: a subset U of B determines an injective map $U \to B$, $u \mapsto u$.

7.2 Relations

The truth of 'n is even' depends on the natural number n. It corresponds to a subset of $\mathbb{N}$: the set of all even natural numbers. The truth of something like $m \leq n$ depends on the natural numbers m and n, or what amounts to the same, it depends on the ordered pair (m,n). Ordered pairs are used to describe *relations*.

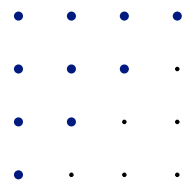


Figure 7.2: The relation $\leq$ of example 7.6

7.6 Example. For elements m and n of $\mathbb{N}_4$ we consider $m \leq n$. It determines a relation in the set $\mathbb{N}_4$. This relation can be given by enumerating all ordered pairs (m, n) satisfying $m \leq n$: $(0,0)$, $(0,1)$, $(0,2)$, $(0,3)$, $(1,1)$, $(1,2)$, $(1,3)$, $(2,2)$, $(2,3)$, $(3,3)$. They all are elements of the set of all ordered pairs (m, n) with $m, n \in \mathbb{N}_4$, that is they form a subset of the Cartesian product $\mathbb{N}_4 \times \mathbb{N}_4$, also denoted by $\mathbb{N}_4^2$, see Figure 7.2. In mathematics one usually identifies relations with such subsets.

7.7 Definition. A *relation* in a set A is a subset of A^2.

If R is a relation in A, then '$(a, b) \in R$' is a good notation. Often the notation '$a\,R\,b$' is used, the so-called *infix* notation.

Since a relation in a set A is just a subset of $A \times A$, it is easy to compute the number of relations in A:

$$\text{number of relations in } A = \#(\mathcal{P}(A \times A)) = 2^{\#(A\times A)} = 2^{\#(A)^2}.$$

So for example: if $\#(A) = 4$, then this number is $2^{4^2} = 2^{16} = 65536$. There are many relations in such a small set!

Special relations

Often relations have nice properties and for such relations we have special names.

7.8 Definition. A relation R in a set A is called *reflexive* if $a\,R\,a$ (that is $(a, a) \in R$) for all $a \in A$.

The relation $\leq$ in $\mathbb{N}$ is reflexive, because every number is (less than or) equal to itself.

In a picture of a relation in a set A (as part of the product set A^2) reflexivity means that it contains the diagonal.

7.9 Definition. A relation R in a set A is called *transitive* if for all $a, b, c \in A$:

$$\text{if } a\,R\,b \text{ and } b\,R\,c, \text{ then } a\,R\,c.$$

The relation $\leq$ in $\mathbb{N}$ is transitive: if $a \leq b$ and $b \leq c$, then $a \leq c$.

In terms of a picture it is not so easy to tell what transitivity means.

7.10 Definition. A relation R in a set A is called *symmetric* if for all $a, b \in A$:

$$\text{if } a\,R\,b, \text{ then } b\,R\,a.$$

The relation $\leq$ in the set $\mathbb{N}$ is not symmetric. We have $0 \leq 1$, but not $1 \leq 0$.

For a picture of a relation symmetry means that it is symmetric under reflection in the diagonal.

7.11 Definition. A relation R in a set A is called *antisymmetric* if for every $a, b \in A$

$$\text{if } a\,R\,b \text{ and } b\,R\,a\,, \text{ then } a = b.$$

The relation $\leq$ in $\mathbb{N}$ is antisymmetric and the same holds for $\geq$.

Under reflection in the diagonal of a picture of the relation off-diagonal elements in the relation map to elements not in the relation.

Two important types of relations are the orderings and the equivalence relations. Here are their definitions:

7.12 Definition. A relation R in a set A is called an *ordering* of A if R is reflexive, antisymmetric and transitive.

The relation $\leq$ in $\mathbb{N}$ is an ordering of $\mathbb{N}$. That is exactly proposition 4.29.

7.13 Definition. A relation is called an *equivalence relation* if it is reflexive, symmetric and transitive.

7.14 Example. Let the relation $\equiv$ in $\mathbb{N}$ be defined as follows:

$$a \equiv b \iff a + b \text{ is even.}$$

This is an equivalence relation. The reflexivity and the symmetry are obvious. The transitivity is demonstrated as follows:

> Suppose a, b and c are natural numbers with $a \equiv b$ and $b \equiv c$. Then $a + b$ and $b + c$ are even, say $a + b = 2k$ and $b + c = 2l$ with $k, l \in \mathbb{N}$. Then $a + 2b + c = 2k + 2l$ and so $a + c = 2k + 2l - 2b = 2(k + l - b)$. From this it follows that $a + c$ is even, that is $a \equiv c$.

Hence for all $a, b, c \in \mathbb{N}$ with $a \equiv b$ and $b \equiv c$ it also holds that $a \equiv c$.

7.15 Example. The relation $\asymp$ in $\mathbb{N}$ defined by

$$a \asymp b \iff ab \neq 0$$

is not reflexive, because $0 \asymp 0$ does not hold. It is the only reason why it is not an equivalence relation. All other requirements are fulfilled.

7.3 Equivalence Relations

Let A be a set. As discussed in section 2.3 subsets of A can be given by a property $P(a)$ which elements a of A might have:

$$\{\, a \in A \mid P(a) \,\}.$$

For a given $U \subseteq A$ one can take for $P(a)$ the property $a \in U$.

As subsets are given by properties, partitions are given by equivalence relations. That is what this section is about. For an equivalence relation we often use a symbol like $\sim$ or $\simeq$ instead of a letter like R. The idea of an equivalence relation is that it expresses that elements in a sense are similar. A symbol like $\sim$ is then more suggestive.

7.16 Example. For the equivalence relation $\equiv$ of example 7.14 it holds that even numbers are similar and so are the odd numbers, whereas an even number and an odd number are not. So the equivalence relation $\equiv$ is connected to the subdivision of $\mathbb{N}$ into two subsets: one of the even numbers and one of the odd numbers. The relation determines this partition of $\mathbb{N}$.

We will show that every equivalence relation determines a partition and visa versa.

7.17 Definition. Let $\sim$ be an equivalence relation in a set A and let $a \in A$. Then the set

$$\{\, x \in A \mid x \sim a \,\}$$

is called the *equivalence class* (with respect to $\sim$) of the element a. It is a subset of A and is denoted by $[a]_\sim$ or $[a]$ for short. The set

$$\{\, [a]_\sim \mid a \in A \,\}$$

of all equivalence classes with respect to $\sim$ we denote by $A/\!\sim$.

7.18 Lemma. *Let $\sim$ be an equivalence relation in a set A. Then*

$$\textit{for all } a, b \in A\text{:} \quad a \sim b \iff [a]_\sim = [b]_\sim.$$

Proof.

$\Rightarrow$: Suppose $x \in [a]_\sim$. Then $x \sim a$. Because $a \sim b$, it follows from the transitivity that also $x \sim b$. So $x \in [b]_\sim$.

So for every $x \in [a]_\sim$ we have $x \in [b]_\sim$, that is $[a]_\sim \subseteq [b]_\sim$. Since the relation is symmetric and so also $b \sim a$, we have similarly that $[b]_\sim \subseteq [a]_\sim$. Hence $[a]_\sim = [b]_\sim$.

$\Leftarrow$: The reflexivity implies that $a \in [a]_\sim$. So $a \in [b]_\sim$, that is $a \sim b$. □

7.19 Proposition. *Let $\sim$ be an equivalence relation in a set A. Then $A/\!\sim$ is a partition of A.*

Proof. Consider the map $f\colon A \to A/\!\sim$ defined by $f(a) = [a]_\sim$. From lemma 7.18 it follows that $A/\!\sim$ is the partition A_f of A. □

7.20 Example. For the equivalence relation $\equiv$ of example 7.14 we have

$$[0] = \{\, x \in \mathbb{N} \mid x \equiv 0 \,\} = \{\, x \in \mathbb{N} \mid x \text{ is even} \,\},$$
$$[1] = \{\, x \in \mathbb{N} \mid x \equiv 1 \,\} = \{\, x \in \mathbb{N} \mid x \text{ is odd} \,\}.$$

So $\mathbb{N}/\!\equiv\ = \{[0], [1]\}$. A system of representatives is $\{0, 1\}$, but there are many others, for example $\{2012, 101\}$.

Let A be a set. For a partition Φ of A there is an equivalence relation $\sim$ such that the equivalence classes with respect to this relation are just the classes of Φ: define $a \sim b$ as $[a]_\Phi = [b]_\Phi$. Conversely for an equivalence relation $\sim$ in A there is a partition $A/\!\sim$ of A which in turn determines the original equivalence relation $\sim$. Partitions and equivalence relations are linked one to one to each other. There are as many partitions in a set as there are equivalence relations in that set.

7.4 Construction of the Integers

Suppose we already have what we want: an extension $\mathbb{Z}$ of $\mathbb{N}$ such that $m + x = n$ always has a solution and the rules of arithmetic hold for this extension as well. Integers we see as differences of natural numbers and so we have a surjective map

$$f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{Z},\ (n, m) \mapsto n - m.$$

This f determines a partition of $\mathbb{N} \times \mathbb{N}$ and the corresponding equivalence relation is

$$(n_1, m_1) \sim (n_2, m_2) \iff f(n_1, m_1) = f(n_2, m_2) \iff n_1 - m_1 = n_2 - m_2.$$

The rules of arithmetic for $\mathbb{Z}$ imply that this equivalence relation can also be described in terms of natural numbers alone:

$$(n_1, m_1) \sim (n_2, m_2) \iff n_1 + m_2 = n_2 + m_1.$$

Thus we have a bijection

$$(\mathbb{N} \times \mathbb{N})/\!\sim\ \to \mathbb{Z}.$$

Conclusion: *if* there exists an extension $\mathbb{Z}$ of $\mathbb{N}$ as desired, *then* the relation $\sim$ in $\mathbb{N} \times \mathbb{N}$ is an equivalence relation and the elements of $\mathbb{Z}$ correspond to equivalence classes in $\mathbb{N} \times \mathbb{N}$. For constructing $\mathbb{Z}$ it is now clear what to do: prove that $\sim$ is an equivalence relation and define $\mathbb{Z}$ to be the set $(\mathbb{N} \times \mathbb{N})/\!\sim$. Next an addition in this set has to be defined, etc. We start with the definition of the relation $\sim$:

7.21 Definition. For $n_1, n_2, m_1, m_2 \in \mathbb{N}$ we define

$$(n_1, m_1) \sim (n_2, m_2) \iff n_1 + m_2 = m_1 + n_2.$$

(Thus the relation $\sim$ is a relation in $\mathbb{N}^2$.)

7.22 Lemma. *$\sim$ is a equivalence relation in $\mathbb{N}^2$.*

Proof. From the definition of $\sim$ it immediately follows that $(n, m) \sim (n, m)$ for all $n, m \in \mathbb{N}$. So reflexivity is clear, as is symmetry. We prove the transitivity.

> Let $n_1, m_1, n_2, m_2, n_3, m_3$ be natural numbers such that $(n_1, m_1) \sim (n_2, m_2)$ and $(n_2, m_2) \sim (n_3, m_3)$. Then $n_1 + m_2 = m_1 + n_2$ and $n_2 + m_3 = m_2 + n_3$, and therefore:
>
> $$n_1 + m_2 + n_2 + m_3 = m_1 + n_2 + m_2 + n_3.$$
>
> By the cancellation law for addition in $\mathbb{N}$,
>
> $$n_1 + m_3 = m_1 + n_3$$
>
> and so $(n_1, m_1) \sim (n_3, m_3)$.

So the relation $\sim$ is transitive. Hence $\sim$ is an equivalence relation. □

7.23 Definition. $\mathbb{Z} = \mathbb{N}^2/\!\sim$. The elements of $\mathbb{Z}$ are called *integers*. The equivalence class $[(n, m)]_\sim$ of $(n, m) \in \mathbb{N}^2$ we denote for the time being as $[n, m]$.

> Think of $[n, m]$ as being the difference of the natural numbers n and m. See Figure 7.3 for a picture of the partition of $\mathbb{N}^2$: connected dots form an equivalence class, an integer that is.

7.4.1 Addition in $\mathbb{Z}$

We will define the addition of integers. Since we want the usual rules of arithmetic to remain valid, we are forced to do it in such a way that:

$$\begin{array}{ccccccc} [n_1, m_1] & + & [n_2, m_2] & = & [n_1 + n_2 & , & m_1 + m_2] \\ & \uparrow & & & \uparrow & & \uparrow \\ & \text{addition in } \mathbb{Z} & & & \text{addition} & & \text{addition} \\ & \text{to be defined} & & & \text{in } \mathbb{N} & & \text{in } \mathbb{N} \end{array}$$

> It has to be this way, because if we see pairs $[n, m]$ as differences $n - m$ and if we insist on the rules of arithmetic being valid also in $\mathbb{Z}$, then
>
> $$(n_1 - m_1) + (n_2 - m_2) = (n_1 + n_2) - (m_1 + m_2).$$

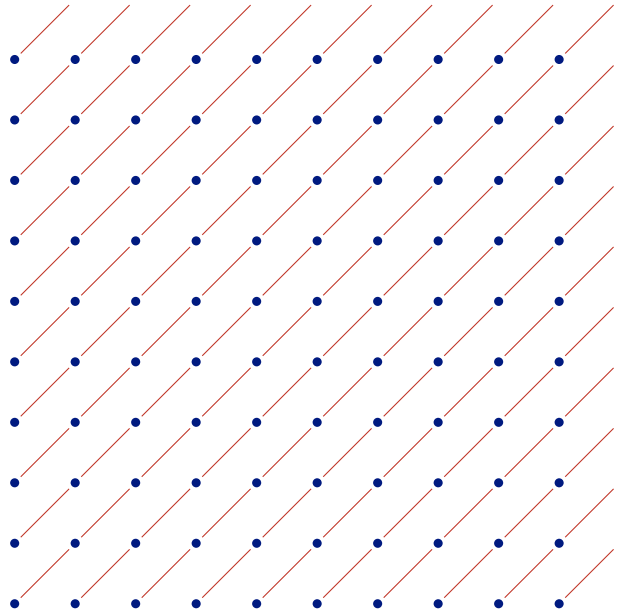


Figure 7.3: The partition $\mathbb{Z}$ of $\mathbb{N}^2$

Therefore, we are forced to define the sum of $a, b \in \mathbb{Z}$ as follows: choose $n_1, m_1 \in \mathbb{N}$ such that $a = [n_1, m_1]$ and choose $n_2, m_2 \in \mathbb{N}$ such that $b = [n_2, m_2]$, then

$$a + b = [n_1 + n_2, m_1 + m_2].$$

A difficulty when defining addition this way is that the result $[n_1 + n_2, m_1 + m_2]$ might depend on the choice of the representatives of the classes a and b. In fact it does not and this follows from the following.

7.24 Lemma. *Le t* $n_1, m_1, n_1', m_1', n_2, m_2, n_2', m_2' \in \mathbb{N}$ *be such that* $(n_1', m_1') \sim (n_1, m_1)$ *and* $(n_2', m_2') \sim (n_2, m_2)$. *Then*

$$(n_1' + n_2', m_1' + m_2') \sim (n_1 + n_2, m_1 + m_2).$$

Proof. This follows directly from definition of $\sim$. □

7.25 Definition. For $a, b \in \mathbb{Z}$ we define the *sum* of a and b as follows:

$$a + b = [n_1 + n_2, m_1 + m_2],$$

if $a = [n_1, m_1]$ and $b = [n_2, m_2]$.

Next we will derive rules for the addition in $\mathbb{Z}$. All we need are the rules of addition in $\mathbb{N}$.

7.26 Proposition. *$\mathbb{Z}$ together with the addition $+$ is an abelian group, that is*

- *The addition is **associative**:*
  $$a + (b + c) = (a + b) + c \quad \textit{(for all } a, b, c \in \mathbb{Z}\textit{)}.$$
- *The addition is **commutative**:*
  $$a + b = b + a \quad \textit{(for all } a, b \in \mathbb{Z}\textit{)}.$$
- $[0, 0]$ *is a **neutral element** (or **zero element**) for the addition:*
  $$a + [0, 0] = a \quad \textit{(for all } a \in \mathbb{Z}\textit{)}.$$
- *Every element a has an **opposite** $-a$:*
  $$a + (-a) = [0, 0].$$

Proof.

Associativity: Choose $m_1, n_1, m_2, n_2, m_3, n_3 \in \mathbb{N}$ such that $a = [n_1, m_1]$, $b = [n_2, m_2]$ and $c = [n_3, m_3]$. Then

$$(a+b)+c = ([n_1+n_2, m_1+m_2])+[n_3, m_3] = [(n_1+n_2)+n_3, (m_1+m_2)+m_3].$$

Similarly
$$a + (b + c) = [n_1 + (n_2 + n_3), m_1 + (m_2 + m_3)].$$
So $(a + b) + c = a + (b + c)$.

Commutativity: Choose $m_1, n_1, m_2, n_2 \in \mathbb{N}$ such that $a = [n_1, m_1]$ and $b = [n_2, m_2]$. Then

$$a + b = [n_1, m_1] + [n_2, m_2] = [n_1 + n_2, m_1 + m_2]$$

and similarly
$$b + a = [n_2 + n_1, m_2 + m_1].$$

Neutral element: Choose $m, n \in \mathbb{N}$ such that $a = [n, m]$. Then

$$a + [0, 0] = [n, m] + [0, 0] = [n + 0, m + 0] = [n, m] = a.$$

Opposite: Choose $m, n \in \mathbb{N}$ such that $a = [n, m]$ and put $b = [m, n]$. Then

$$a + b = [n, m] + [m, n] = [n + m, m + n] = [0, 0]. \qquad \square$$

Note how easily these rules for $\mathbb{Z}$ are derived from the rules for $\mathbb{N}$. The existence of opposites is new. Equations $a + x = b$ are not always solvable in $\mathbb{N}$, but in $\mathbb{Z}$ they are: the (unique) solution is $x = b - a$, that is $x = b + (-a)$.

In $\mathbb{N}$ we had a cancellation law for addition. This law holds in $\mathbb{Z}$ as well. For a proof only the rules for addition in $\mathbb{Z}$ are needed: if $a + b = c + b$, then (writing 0 for the zero element $[0, 0]$):

$$\begin{aligned} a &= a + 0 = a + (b + (-b)) = (a + b) + (-b) \\ &= (c + b) + (-b) = c + (b + (-b)) = c + 0 = c. \end{aligned}$$

7.4.2 $\mathbb{N}$ as part of $\mathbb{Z}$

There still is a little problem. The set $\mathbb{N}$ is not a subset of the set $\mathbb{Z}$. However, inside $\mathbb{Z}$ there is a subset which with the addition of $\mathbb{Z}$ is very much like $\mathbb{N}$ itself:

$$\underline{\mathbb{N}} = \{\, [n, 0] \mid n \in \mathbb{N} \,\}.$$

This set can be seen as a copy of $\mathbb{N}$:

$$[m, 0] = [n, 0] \iff m = n$$

and

$$[m, 0] + [n, 0] = [m + n, 0].$$

So addition in $\underline{\mathbb{N}}$ amounts to the same as addition in $\mathbb{N}$. Writing n for $[n, 0]$ and so $-n$ for $-[n, 0]$, it looks like as if $\mathbb{N} \subseteq \mathbb{Z}$ and everything is as we want it to be. For the integer $[n, m]$ we then have

$$[n, m] = [n, 0] + [0, m] = [n, 0] - [m, 0] = n - m.$$

We postpone this change in notation till after the introduction of the multiplication in $\mathbb{Z}$.

Note that the set $\underline{\mathbb{N}}$, together with the element $[0, 0]$ and the transformation $\underline{\sigma}$ with $\underline{\sigma}([n, 0]) = [n + 1, 0]$, satisfies Peano's axioms. So also for that reason the set $\underline{\mathbb{N}}$ can be seen as the system of natural numbers.

In fact we have an injective map $i \colon \mathbb{N} \to \mathbb{Z},\ n \mapsto [n, 0]$ and it satisfies $i(\sigma(n)) = \underline{\sigma}(i(n))$ for all $n \in \mathbb{N}$. The image $i_*(\mathbb{N})$ is the set $\underline{\mathbb{N}}$.

7.4.3 Multiplication in $\mathbb{Z}$

By now we have the set $\mathbb{Z}$ together with an addition satisfying the familiar rules. Next we will define a multiplication in $\mathbb{Z}$. Since we see the integers as differences of natural numbers and we want the rules of arithmetic to remain valid, for the definition of the product there is no choice: $(n_1 - m_1) \cdot (n_2 - m_2) = n_1 n_2 + m_1 m_2 - (m_1 n_2 + n_1 m_2)$.

7.27 Definition. Let a and b be integers. We define the *product* $a \cdot b$ of a and b as follows. Choose $m_1, n_1, m_2, n_2 \in \mathbb{N}$ such that $a = [n_1, m_1]$ and $b = [n_2, m_2]$. Then

$$a \cdot b = [n_1 n_2 + m_1 m_2, m_1 n_2 + n_1 m_2].$$

Also in this case the verification that the definition does not depend on the choices made is straightforward:

7.28 Lemma. *Let $n_1, m_1, n_2, m_2, n_1', m_1', n_2', m_2' \in \mathbb{N}$ such that $(n_1', m_1') \sim (n_1, m_1)$ and $(n_2', m_2') \sim (n_2, m_2)$. Then*

$$(n_1' n_2' + m_1' m_2', m_1' n_2' + n_1' m_2') \sim (n_1 n_2 + m_1 m_2, m_1 n_2 + n_1 m_2).$$

Proof. The proof is in two steps.

First assume that $(n_2', m_2') = (n_2, m_2)$. We have

$$\begin{aligned} n_1' n_2 + m_1' m_2 + m_1 n_2 + n_1 m_2 &= (n_1' + m_1) n_2 + (m_1' + n_1) m_2 \\ &= (m_1' + n_1) n_2 + (n_1' + m_1) m_2 = m_1' n_2 + n_1' m_2 + n_1 n_2 + m_1 m_2. \end{aligned}$$

Hence

$$(n_1' n_2 + m_1' m_2, m_1' n_2 + n_1' m_2) \sim (n_1 n_2 + m_1 m_2, m_1 n_2 + n_1 m_2). \tag{7.2}$$

Now assume that $(n_1, m_1) = (n_1', m_1')$. Then

$$\begin{aligned} n_1' n_2' + m_1' m_2' + m_1' n_2 + n_1' m_2 &= (n_1' + m_1') n_2' + (m_1' + n_1') m_2' \\ &= (m_1' + n_1') n_2' + (n_1' + m_1') m_2' = m_1' n_2 + n_1' m_2 + n_1' n_2' + m_1' m_2. \end{aligned}$$

Hence

$$(n_1' n_2' + m_1' m_2', m_1' n_2' + n_1' m_2') \sim (n_1' n_2 + m_1' m_2, m_1' n_2 + n_1' m_2). \tag{7.3}$$

By the transitivity of $\sim$ the lemma follows from (7.2) and (7.3). □

7.29 Proposition. *The set $\mathbb{Z}$ together with the operations $+$ and $\cdot$ as defined above is a commutative ring, i.e. $\mathbb{Z}$ together with the addition is an abelian group and the following rules hold for all $a, b, c \in \mathbb{Z}$:*

- *The multiplication is* ***associative****:*
  $a \cdot (b \cdot c) = (a \cdot b) \cdot c.$
- *The multiplication is* ***commutative****:*
  $a \cdot b = b \cdot a.$
- $[1, 0]$ *is a* ***neutral element*** *(or* ***unity element****) for the multiplication:*
  $a \cdot [1, 0] = a.$
- *The multiplication is* ***distributive*** *over the addition:*
  $a \cdot (b + c) = a \cdot b + a \cdot c.$

Proof. Put $a = [n_1, m_1]$, $b = [n_2, m_2]$ and $c = [n_3, m_3]$.

Associativity: $a \cdot (b \cdot c)$ and $(a \cdot b) \cdot c$ are equal to respectively

$$[n_1(n_2n_3+m_2m_3)+m_1(m_2n_3+n_2m_3), m_1(m_2n_3+n_2m_3)+n_1(m_2n_3+n_2m_3)]$$

and

$$[(n_1n_2+m_1m_2)n_3+(m_1n_2+n_1m_2)m_3, (m_1n_2+n_1m_2)n_3+(n_1n_2+m_1m_2)m_3].$$

Commutativity: $a \cdot b$ and $b \cdot a$ are equal to respectively

$$[n_1n_2 + m_1m_2, m_1n_2 + n_1m_2] \quad \text{and} \quad [n_2n_1 + m_2m_1, m_2n_1 + m_1n_2].$$

Unity element: $[1, 0]$ is the unity element: $a \cdot [1, 0] = [n_1 + 0, 0 + m_1] = a$.

Distributivity: $a \cdot (b + c)$ and $a \cdot b + a \cdot c$ are equal to respectively

$$[n_1(n_2 + n_3) + m_1(m_2 + m_3), m_1(n_2 + n_3) + n_1(m_2 + m_3)]$$

and

$$[(n_1n_2 + m_1m_2) + (n_1n_3 + m_1m_3), (m_1n_2 + n_1m_2) + (m_1n_3 + n_1m_3)]. \quad \square$$

7.4.4 Exponentiation in $\mathbb{Z}$

Just as in $\mathbb{N}$, exponentiation in $\mathbb{Z}$ is repeated multiplication.

7.30 Definition. Let $a \in \mathbb{Z}$. For natural numbers $n \in \mathbb{N}$ the integers a^n are defined by

$$\begin{cases} a^0 = 1, \\ a^{n+1} = aa^n \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Note that we only take nth powers for $n \in \mathbb{N}$. This definition of exponentiation can be given for any monoid. Then the usual rules for exponentiation hold. In $\mathbb{Z}$ the exponentiation is an extension of the exponentiation in $\mathbb{N}$. For $-a$ with $a \in \mathbb{N}$ we now have $(-a)^n = (-1)^n a^n$.

7.4.5 Ordering of $\mathbb{Z}$

7.31 Definition. The ordering $\leq$ of $\mathbb{N}$ can be extended to an ordering $\leq$ of $\mathbb{Z}$:

$$a \leq b \iff b - a \in \mathbb{N}.$$

Here we see $\mathbb{N}$ as a part of $\mathbb{Z}$ and that is what we will do from now on. Instead of $a \leq b$ we also write $b \geq a$. We have: $a \in \mathbb{N} \iff a \geq 0$.

For $\leq$ the requirements for an ordering hold and also the rules that connect the ordering with addition and multiplication:

7.32 Proposition. *The relation $\leq$ is an ordering of $\mathbb{Z}$, i.e. it is **reflexive**, **antisymmetric** and **transitive**. Moreover for all $a, b, c \in \mathbb{Z}$:*

- *if $a \leq b$, then $a + c \leq b + c$,*
- *if $a \leq b$ and $c \geq 0$, then $ac \leq bc$.*

Proof.

Reflexivity: $a \leq a$, because $a - a = 0 \in \mathbb{N}$.

Antisymmetry: If $b - a \in \mathbb{N}$ and $a - b \in \mathbb{N}$, then, because $(a - b) + (b - a) = 0$, also $a - b = 0$.

Transitivity: If $b - a \in \mathbb{N}$ and $c - b \in \mathbb{N}$, then also $c - a = (c - b) + (b - a) \in \mathbb{N}$.

Relation with addition: If $b - a \in \mathbb{N}$, the also $(b + c) - (a + c) \in \mathbb{N}$.

Relation with multiplication: If $b - a \in \mathbb{N}$ and $c \in \mathbb{N}$, then also $bc - ac = (b - a)c \in \mathbb{N}$. □

Other rules can be derived. For instance, if in the last rule $c \geq 0$ is replaced by $c \leq 0$, then $-c \geq 0$ and so by the same rule $ac - bc = (b - a)(-c) \in \mathbb{N}$, that is $ac \geq bc$.

7.4.6 Absolute value

7.33 Definition. Let $a \in \mathbb{Z}$. We define the *absolute value* of a, notation $|a|$, as follows:

$$|a| = \begin{cases} a & \text{if } a \geq 0, \\ -a & \text{otherwise.} \end{cases}$$

Thus we have $a = \pm|a|$. Integers are introduced as 'formal' differences of natural numbers: if $a = [n, m]$, then—with the identification of natural numbers n with the integers $[n, 0]$—we have $a = n - m$. From the definition of $|a|$ it follows directly that $|a| \geq 0$. The absolute value is a map $\mathbb{Z} \to \mathbb{N}$, $a \mapsto |a|$.

This absolute value has properties we will require more generally for absolute values:

7.34 Proposition. *The absolute value $\mathbb{Z} \to \mathbb{N},\ a \mapsto |a|$ has the following properties:*

(i) $|a| = 0 \iff a = 0$ *(for all $a \in \mathbb{Z}$),*
(ii) $|ab| = |a| \cdot |b|$ *(for all $a, b \in \mathbb{Z}$),*
(iii) $|a + b| \leq |a| + |b|$ *(for all $a, b \in \mathbb{Z}$).*

Proof.

(i) Follows directly from the definition.
(ii) $|a| \cdot |b| = (\pm a)(\pm b) = \pm ab = \pm|ab|$, and since $|ab|, |a| \cdot |b| \geq 0$, we have $|ab| = |a| \cdot |b|$.
(iii) $a \leq |a|$ and $b \leq |b|$, so $a + b \leq |a| + |b|$. Because $|-a| = |a|$, we have similarly $-a - b \leq |a| + |b|$. Hence: $|a + b| = \pm(a + b) \leq |a| + |b|$. □

7.35 Proposition. *The commutative ring $\mathbb{Z}$ has no zero divisors, i.e.*

for all $a, b \in \mathbb{Z}$ we have: if $ab = 0$, then $a = 0$ or $b = 0$.

Proof. Let a and b be integers with $ab = 0$. Because also $|a|.|b| = 0$, lemma 4.21 implies that $|a| = 0$ or $|b| = 0$. □

It follows that the cancellation law for multiplication also holds for the multiplication of integers.

7.36 Corollary. *Let a, b and c be integers. Then:*

if $ac = bc$ and $c \neq 0$, then $a = b$.

Proof. If $ac = bc$, then $(a - b)c = 0$, and if moreover $c \neq 0$, then $a - b = 0$. □

7.5 Algebraic Structures

An algebraic structure consists of a set and a number of operations in that set which satisfy some given properties, usually called axioms. One example we have already met is the monoid: a set with an associative operation for which there is a neutral element. In this chapter we came across some algebraic structures which are important for modern mathematics.

The idea of Euclid's axioms for geometry, for Peano's axioms for the natural numbers and for the axioms of set theory is that they are *categorical*, which means that the structure they describe is in a sense completely determined. These axioms are meant to be a foundation for mathematics. With axioms for an algebraic structure the situation is quite different. There the power of the approach lies in the abundance of examples.

7.5.1 Groups and abelian groups

The notion of (abelian) group is abstract. We will give the definition. It enables us to give compact formulations for theorems. Group theory is a highly developed part of algebra. In this book we will not go into this theory and confine ourselves mainly to the use of its terminology.

7.37 Definition. An *abelian group* is a set A together with an operation $A^2 \to A$, $(a, b) \mapsto a + b$ such that a number of requirements (axioms) are fulfilled:

(G1) Associativity: $(a + b) + c = a + (b + c)$ for all $a, b, c \in A$.

(G2) Neutral element: There is an $n \in A$ such that $a + n = n + a = a$ for all $a \in A$.

(G3) Opposite: For all $a \in A$ there exists a $b \in A$ such that $a+b = b+a = n$ (with n as in (G2)).

(G4) Commutativity: $a + b = b + a$ for all $a, b \in A$.

If n' is also a neutral element, then by (G2): $n' = n + n' = n$. So there is a unique neutral element. That element is usually denoted by 0 and is called *zero* or the *zero element*.

If there are for a given a elements b and b' such that $a + b = a + b' = 0$, then $b' = b' + 0 = b' + (a + b) = (b' + a) + b = 0 + b = b$. We call b the *opposite* of a and denote it by $-a$. Thus: $-a$ is the unique element such that $a+(-a) = (-a)+a = 0$.

We have seen that for the addition in $\mathbb{Z}$ the cancellation law holds. For the proof we only used the fact that $\mathbb{Z}$ together with the operation $+$ is an abelian group. Hence, we actually derived that the cancellation law holds in any abelian group. That is the advantage of abstraction: if something is an abelian group, then anything that is derived from just the group axioms holds for that object in particular.

Without axiom (G4), the commutativity, we have the definition of the abstract notion of *group*. (That is why in the formulation of the axioms (G2) and (G3) no use is made of the commutativity.) The operation in a group usually is not seen as addition (with a $+$), but as a multiplication (with another notation for the operation: $(a, b) \mapsto ab$), especially when commutativity does not hold. The neutral element is then denoted by 1 and is called *one* or the *unity element*. Also the term 'opposite' is not used in that case, but the term is *inverse* and the notation is: a^{-1}. Thus we have:

7.38 Definition. A *group* is a set G together with an operation $G^2 \to G$, $(g, h) \mapsto gh$ such that the following holds:

(G1) Associativity: $(gh)k = g(hk)$ for all $g, h, k \in G$.

(G2) Neutral element: There is a $1 \in G$ such that $g1 = 1g = g$ for all $g \in G$.

(G3) Inverse: For all $g \in G$ there is an $h \in G$ such that $gh = hg = 1$ (with 1 as in (G2)).

7.5.2 Rings, commutative rings, integral domains

A structure with operations addition and multiplication which satisfy the usual rules of arithmetic is called a ring. The exact definition is as follows.

7.39 Definition. A *ring* is a set R together with two operations, an 'addition' $R^2 \to R, (a, b) \mapsto a + b$ and a 'multiplication' $R^2 \to R$, $(a, b) \mapsto ab$ (or: $a \cdot b$), such that R together with $+$ is an abelian group and moreover the following axioms are satisfied:

(R1) Associativity of the multiplication:
$(ab)c = a(bc)$ for all $a, b, c \in R$.

(R2) Unity element: there is a $1 \in R$ such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in R$.

(R3) Distributivity of the multiplication over the addition:
$a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ for all $a, b, c \in R$.

If also the following axiom is satisfied

(R4) Commutativity of the multiplication: $ab = ba$ for all $a, b \in R$,

then R is called a *commutative ring*.

In a ring it is possible that $ab = 0$ while $a \neq 0$ and $b \neq 0$. For commutative rings we have a name for this.

7.40 Definition. Let R be a commutative ring. An element $a \neq 0$ for which there is an element $b \neq 0$ such that $ab = 0$ is called a *zero divisor* of the ring R. If a commutative ring R has no zero divisors, then R is called an *integral domain* or simply a *ring without zero divisors*.

So the ring $\mathbb{Z}$ is an example of an integral domain.

7.6 Orderings

We constructed the ring $\mathbb{Z}$ using an equivalence relation. That is why we took a closer look at equivalence relations. In $\mathbb{N}$ and also in $\mathbb{Z}$ we have the ordering $\leq$. An ordering is a reflexive, antisymmetric, transitive relation, see definition 7.12. In this section we consider this other important type of relations. For a given, but arbitrary, ordering we will often use the symbol $\preceq$: it looks like $\leq$, but that symbol usually has a more specific meaning.

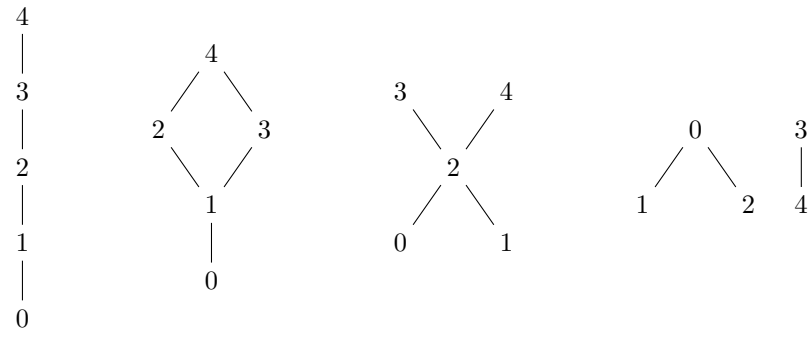


Figure 7.4: Four orderings of $\mathbb{N}_5$

7.41 Definition. Let $\preceq$ be an ordering of a set A. Then $(A, \preceq)$ is called an *ordered* set.

7.42 Examples. $(\mathbb{N}, \leq)$, $(\mathbb{N}, \geq)$, $(\mathbb{Z}, \leq)$ and $(\mathbb{Z}, \geq)$ are examples of ordered sets. Note that the relation $\geq$ is an ordering as well. If $\preceq$ is an ordering, then so is $\succeq$, defined by $a \succeq b \iff b \preceq a$. If $(A, \preceq)$ is an ordered set and U is a subset of A, then the restriction of $\preceq$ to U is an ordering of U.

Orderings of sets which are not too big can be represented by a picture. Figure 7.4 for example consists of pictures of four different orderings of the set $\mathbb{N}_5$. The idea is that if you can move along the connections from a to b in an upward direction, then this means that $a \preceq b$. Such a picture is called a *Hasse diagram* of the ordering.

We will distinguish some types of orderings.

7.43 Definition. An ordering $\preceq$ of a set A is called *total* if for every $a, b \in A$ the following holds: $a \preceq b$ or $b \preceq a$. We then say that $(A, \preceq)$ is a *totally* ordered set.

7.44 Examples. The orderings $\leq$ and $\geq$ of the sets $\mathbb{N}$ and $\mathbb{Z}$ are total. For $\mathbb{Z}$ this is a simple consequence of the ordering of $\mathbb{N}$ being total.

7.45 Definition. Let $\preceq$ be an ordering of a set A and let U be a subset of A. Then $a \in A$ is called the *least* element of U (with respect to the ordering $\preceq$) if

a) $a \in U$,
b) $a \preceq u$ for all $u \in U$.

If a is the least element of U, then we denote this by $a = \min(U)$. If a is the least element of U with respect to the ordering $\succeq$ (where $a \succeq b \iff b \preceq a$), then a is called the *greatest* element of U (with respect to the ordering $\preceq$), notation $a = \max(U)$.

Helmut Hasse (Kassel 1898 – Ahrensburg 1979)

Hasse was a German mathematician who did a lot of fundamental research in algebraic number theory. Since in that part of mathematics often substructures of a given structure are considered, it is convenient to make diagrams of such collections of objects. That type of diagrams one started to call Hasse diagrams.

If U has a least element, then that element is unique. This follows easily from the antisymmetry of $\preceq$. It depends on U and $\preceq$ whether such a least element exists.

The subset $\mathbb{Z}$ of $\mathbb{Z}$ has no least element with respect to $\leq$.

The subset $\mathbb{N}$ of $\mathbb{N}$ has no greatest element with respect to $\leq$.

7.46 Definition. An ordering $\preceq$ of a set A is called a *well-ordering* if every non-empty subset of A has a least element with respect to $\preceq$. Then $(A, \preceq)$ is called a *well-ordered* set.

We will prove that $(\mathbb{N}, \leq)$ is a well-ordered set. That is so obvious that you might wonder why a proof is needed. As a matter of fact the same holds for mathematical induction, a property of $\mathbb{N}$ we took as an axiom. In a sense the well-ordering of $\mathbb{N}$ is equivalent to mathematical induction. First we will derive a new version of mathematical induction.

7.47 Proposition. *Let U be a subset of $\mathbb{N}$. Suppose that $n \in U$ for every $n \in \mathbb{N}$ which satisfies $\mathbb{N}_n \subseteq U$. Then $U = \mathbb{N}$.*

Proof. By mathematical induction we prove that $\mathbb{N}_n \subseteq U$ for all $n \in \mathbb{N}$. For $n = 0$ this is clear: $\mathbb{N}_0 = \emptyset$.

Suppose $\mathbb{N}_n \subseteq U$ for an $n \in \mathbb{N}$. Then $n \in U$, and so $\mathbb{N}_{n+1} = \mathbb{N}_n \cup \{n\} \subseteq U$.

Hence $\mathbb{N}_n \subseteq U$ for all $n \in \mathbb{N}$. For every $n \in \mathbb{N}$ we have $n \in \mathbb{N}_{n+1} \subseteq U$ and so $\mathbb{N} \subseteq U$. □

7.48 Theorem. *The ordering $\leq$ of $\mathbb{N}$ is a well-ordering.*

Proof. Let U be a subset of $\mathbb{N}$ without a least element. We will show that $U = \emptyset$. Consider $V = \mathbb{N} \setminus U$. To prove that $V = \mathbb{N}$.

Suppose $n \in \mathbb{N}$ with $\mathbb{N}_n \subseteq V$. Then $k \notin U$ for all $k < n$. If $n \in U$, then n would have been the least element of U. So $n \notin U$, that is $n \in V$.

From proposition 7.47 it follows that $V = \mathbb{N}$. □

Proposition 7.47 and theorem 7.48 are variations on the principle of mathematical induction and are often used in proofs. Proposition 7.47 gives the *strong* principle of mathematical induction.

mathematical induction (strong)
Suppose $n \in \mathbb{N}$ such that $P(k)$ for all $k < n$.
...
So $P(n)$.
Hence $P(n)$ for all $n \in \mathbb{N}$ such that $P(k)$ for all $k < n$.
By mathematical induction it follows that $P(n)$ for all $n \in \mathbb{N}$.

With ordinary induction one tries to prove $P(n+1)$ under the assumption $P(n)$. With this strong form of induction one tries to give a proof of $P(n)$ under the assumption that $P(k)$ holds for *all* $k < n$. It might give the impression that forget the case $n = 0$. However. for $n = 0$ it says that $P(0)$ should hold if $P(k)$ holds for all natural numbers $k < 0$, but such natural numbers do not exist. In a proof along the lines of the strong induction principle usually case distinction occurs: cases in which the induction hypothesis is used and cases in which it is not used. In case $n = 0$ it certainly can not be used.

The ordering $\leq$ of $\mathbb{Z}$ is not a well-ordering. Nonempty subsets have a least element only under an extra condition.

7.49 Definition. Let $(A, \preceq)$ be an ordered set. An $a \in A$ is called a *lower bound* of a subset U of A if $a \preceq u$ for all $u \in U$. A subset U is called *bounded below* if there is a lower bound for U.

A $b \in A$ is called an *upper bound* of a subset U of A if $u \preceq b$ for all $u \in U$. The subset U is called *bounded above* if there is an upper bound for U.

7.50 Theorem. *Nonempty subsets of $\mathbb{Z}$ which are bounded below have a least element. Nonempty subsets of $\mathbb{Z}$ which are bounded above have a greatest element.*

Talking about the ordered set $\mathbb{Z}$ we mean $(\mathbb{Z}, \leq)$. A greatest element is a least element with respect to $\geq$.

Proof. If $a \in \mathbb{Z}$ is a lower bound of a nonempty $U \subseteq \mathbb{Z}$, then $-a+U = \{\, -a+u \mid u \in U \,\}$ is a nonempty subset of $\mathbb{N}$. If $c \in \mathbb{N}$ is the least element of $-a+U$, then $a+c$ is the least element of U.

If $b \in \mathbb{Z}$ is an upper bound of a nonempty $V \subseteq \mathbb{Z}$, then $-b$ is a lower bound of $-U = \{\, -u \mid u \in U \,\}$. If d is the least element of $-U$, then $-d$ is the greatest of U. □

A total ordering is not necessarily a well-ordering. The converse however holds.

7.51 Theorem. *Let $\preceq$ be a well-ordering of a set A. Then $\preceq$ is a total ordering.*

Proof.

> Let a, b be elements of A. Then $\{a, b\}$ is a subset of A. Since $\preceq$ is a well-ordering, this subset has a least element. If a is the least, then $a \preceq b$. If b is the least, then $b \preceq a$. So $a \preceq b$ or $b \preceq a$.

Hence for all $a, b \in A$ we have: $a \preceq b$ or $b \preceq a$. This means that the ordering $\preceq$ is total. □

An important example of an ordering is the relation 'is a subset of'. This is a relation in the power set of a given set.

7.52 Example. Let B be a set. The power set $\mathcal{P}(B)$ is an ordered set: $(\mathcal{P}(B), \subseteq)$. Here the inclusion is the ordering. The set $\mathcal{P}(\{1, 2, 3\})$ has 8 elements and Figure 7.5 is a Hasse diagram of this ordered set. This ordering is neither a total ordering nor a well-ordering.

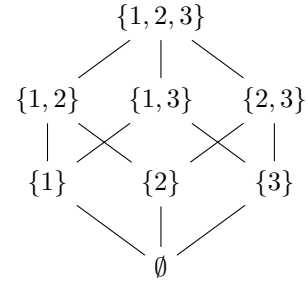


Figure 7.5: Hasse diagram of $\mathcal{P}(\underline{3}, \subseteq)$

7.7 Directed Graphs

In some cases one wants the edges of a graph to have a direction. In the sliding puzzle with 15 blocks on a 4 by 4 board there is no need for a direction, since the sliding goes both ways. The same holds for the Tower of Hanoi. For the well known *solitary* puzzle the situation is different: with each move a piece is taken and the goal is to have only one piece left. If you want to represent this by a graph, then the possible positions of the puzzle can be taken as vertices and the moves as edges with a direction, that is the edges connect a first to a second vertex. The abstract notion is very simple:

7.53 Definition. A *directed graph* (V, E) is a finite set V together with a subset E of V^2. The elements of V are called the *vertices* and the elements of E the *edges* of the directed graph. An edge (v_1, v_2) has a *head* and a *tail*: the head is the vertex v_2, and the tail the vertex v_1.

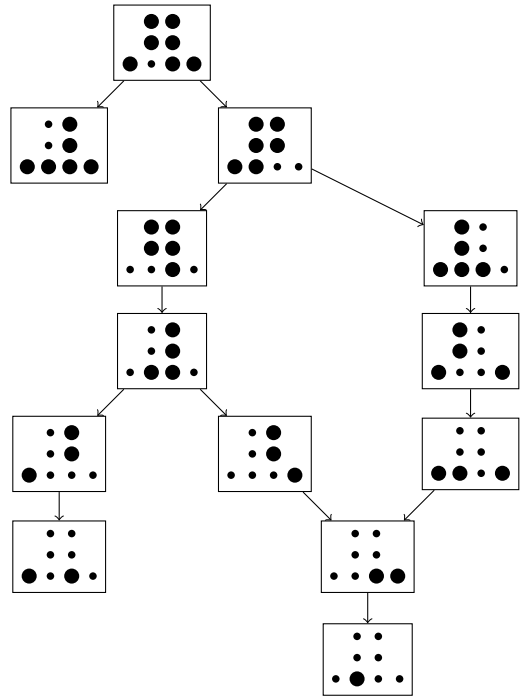
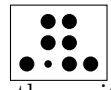


Figure 7.6: Directed graph of a simple solitary puzzle

Thus a directed graph is just a relation in a finite set. The use of the word graph is merely an invitation for a special geometrical interpretation of the relation.

7.54 Example. Figure 7.6 is a picture of a directed graph representing an extremely simplified solitary puzzle: the puzzle with the board  and with the initial position  . The number of possible positions is $2^8 = 256$; the picture contains only the positions which can be reached departing from the initial position.

7.55 Example. In the set $\{1, 2\}$ there are $2^{2^2} = 16$ relations. Each of these relations can be seen as a directed graph with 1 and 2 as vertices. The number of edges is at most $2^2 = 4$. In Figure 7.7 all possible relations are represented (as subsets of ) together with the corresponding directed graph.

For each $R \subseteq A^2$ we have a directed graph (A, R). If R is moreover the graph of a transformation of A, then the picture of this transformation is the same as the picture of the directed graph. It is the special case in which there is for each

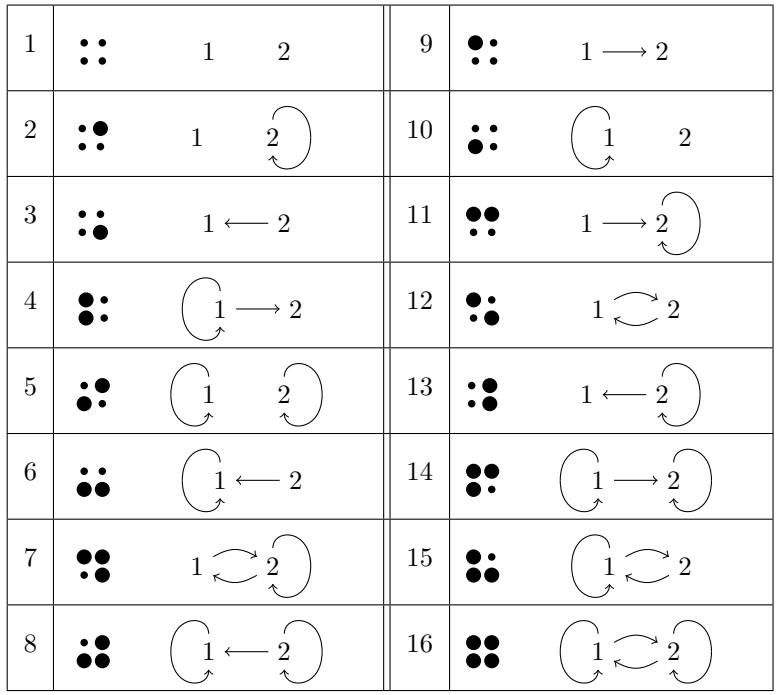


Figure 7.7: Directed graphs having 1 and 2 as vertices

$a \in A$ exactly one edge of type (a, b). (Note that we used the word graph here with two different meanings: the graph as a combinatorial structure and in the sense of graph of a map.)

If R is an ordering, we prefer a different kind of picture, namely the Hasse diagram: since the direction of the arrows is always upwards we do not draw the arrow tips and we also do not draw edges, the existence of which is ensured by transitivity or reflexivity.

If R is an equivalence relation, then we do not draw the edges, but only indicate the equivalence classes. Inside an equivalence class all ordered pairs are edges.

Exercises

1. Which of the following properties are satisfied by the relation R in the set A: reflexivity, symmetry, transitivity, antisymmetry?
   (i) any A and: $a \, R \, b \iff a = b$.

(ii) any A and: $a\,R\,b$ for all $a, b \in A$.

(iii) $A = \mathbb{N}$ and: $a\,R\,b \iff a < b$.

(iv) $A = \mathbb{N}$ and: $a\,R\,b \iff a + b$ is a multiple of 3.

Which of these relations is an ordering and which an equivalence relation?

2. Which of the 16 relations in $\{1, 2\}$ is an ordering, which is an equivalence relation and which is the graph of a transformation? See example 7.55 and Figure 7.7.

3. Let A be a set of 10 elements. Determine the number of
   (i) relations in A,
   (ii) reflexive relations in A,
   (iii) symmetric relations in A,
   (iv) reflexive symmetric relations in A,
   (v) reflexive antisymmetric relations in A.

4. Let R be both an ordering of a set A and an equivalence relation in A. What is R?

5. What is wrong in the following 'derivation' of the proposition: *symmetric, transitive relations are reflexive.*
   Let $\sim$ be a symmetric, transitive relation in a set A.

   > Let a be an element of A. For $b \in A$ such that $a \sim b$ also $b \sim a$ holds because of symmetry of $\sim$. From transitivity it then follows that $a \sim a$.

   Hence $a \sim a$ for all $a \in A$. So $\sim$ is reflexive.

6. Let $\sim$ be a reflexive relation in a set A which also satisfies: if $a \sim b$ and $c \sim b$, then $a \sim c$ (for all $a, b, c \in A$). Is $\sim$ an equivalence relation?

7. How many partitions are there of the set $\{1, 2, 3, 4, 5\}$? And how many equivalence relations are there in this set?

8. Show that the multiplication in $\mathbb{Z}$ is well defined, meaning that it does not depend on the choice of representatives.

9. Is there an equivalence relation in $\mathbb{N}$ with all its equivalence classes infinite and also the equivalence classes infinite in number? If no, give a proof. If yes, give an example.

10. We have constructed $\mathbb{Z}$ as a partition of $\mathbb{N}^2$ which was obtained by an equivalence relation. If we apply the 'same' construction to $\mathbb{Z}$ instead of $\mathbb{N}$, what do we get?

11. From the exercises 13 and 14 of chapter 5 it follows that $\mathbb{Z}$ is countable. How?

12. The sequence (a_n) in $\mathbb{Z}$ is defined by

$$\begin{cases} a_0 = 0, \\ a_{n+1} = a_n + (-1)^n(n+1) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

    Prove that the map $\mathbb{N} \to \mathbb{Z},\ n \mapsto a_n$ is bijective.

13. Let a and b be integers. Prove the following rules

$$a^2 - b^2 = (a - b)(a + b)$$
$$a^3 - b^3 = (a - b)(a^2 + ab + b^2).$$

Indicate which rules of arithmetic have been used.

14. Let a and b be integers with $a^2 = b^2$. Prove that $a = b$ or $a = -b$.

15. Prove that $a^2 + ab + b^2 \geq 0$ for all $a, b \in \mathbb{Z}$ and that $a^2 + ab + b^2 = 0$ only if $a = b = 0$.

16. Let X be a finite nonempty set. For $A, B \in \mathcal{P}(X)$ we define

$$A \sim B \iff \#(A \div B) \text{ is even.}$$

(i) Prove that $\sim$ is an equivalence relation in $\mathcal{P}(X)$.

(ii) How many equivalence classes are there?

17. Let $(A, \preceq)$ be an ordered set and let f be the following map:

$$f\colon A \to \mathcal{P}(A), \quad a \mapsto \{\, x \in A \mid x \preceq a \,\}.$$

(i) Prove that f is injective.

(ii) Prove that f is not surjective.

18. Let A be a set and let B be s subset of A. In $\mathcal{P}(A)$, the power set of A, the relation $\doteq$ is defined by

$$U \doteq V \iff U \cup B = V \cup B \quad (\text{for all } U, V \subseteq A).$$

(i) Show that $\doteq$ is an equivalence relation.

(ii) Let $U \subseteq A$. Assume that B is nonempty. Show that the equivalence class of U has more than one element.

(iii) Prove that

$$\{\, U \in \mathcal{P}(A) \mid U \supseteq B \,\}$$

is a system of representatives of $\mathcal{P}(A)/\doteq$.

19. Show that the number of orderings of a finite set is odd.

8 Numeral Systems

We denote natural numbers by words in the digits 0 up to 9. This is a very rich notation, but so far we have made no use of this. We have described operations in terms of the successor, and these descriptions have little to do with the notation used. For example from the definition of addition it is not immediately clear that something like $6535535301 + 100000 = 6535635301$ holds. As everyone knows there is a close relationship between this notation and the operations of addition, multiplication and exponentiation. As soon as that is clear the application of the rules of arithmetic leads to faster ways of performing the operations, in fact it is the way you have learned at primary school.

The use of ten digits is arbitrary, we could have used any number of digits greater than 1. The purpose here is that an understanding of this kind of notation is reached which is good enough to understand how to convert in a fast way from one notation to another.

Fundamental for the way we denote natural numbers is the notion of 'division with remainder'. In fact many properties of natural numbers are consequences of this division with remainder.

8.1 Division with Remainder

Inside $\mathbb{Z}$ division is not possible in all cases, but what we do have is 'division with remainder'.

8.1 Theorem (Division with remainder). *Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}^+$. Then there are unique $q \in \mathbb{Z}$ and $r \in \mathbb{N}_b$ such that $a = qb + r$.*

Proof. Consider the subset $U = \{\, xb \mid x \in \mathbb{Z} \text{ and } xb \leq a \,\}$ of $\mathbb{Z}$. The integer a is an upper bound for this set. Let m be the greatest number in U. Then $m = qb$ for a $q \in \mathbb{Z}$. Since qb is the greatest in U, it follows that $(q+1)b > a$. So $qb \leq a < qb+b$, that is $0 \leq a - qb < b$. So $a = qb + r$ with $r \in \mathbb{N}_b$.

If also $a = q'b + r'$ with $r' \in \mathbb{N}_b$, then $q'b \leq a$ and so $q'b \leq qb$, because qb was the greatest in U. Since $(q' + 1)b = a + (b - r') > a$ we also have $(q' + 1)b > qb$. So $q' \leq q$ and $q' \geq q$, that is $q' = q$. Then also $r' = r$. □

8.2 Definition. Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}^+$. The unique $q \in \mathbb{Z}$ and $r \in \mathbb{N}_b$ such that $a = qb + r$ are called respectively the *quotient* of a by b and the *remainder* of a after division by b. The number a is called the *dividend*.

8.3 Notations. Division with remainder by a $b \in \mathbb{N}^+$ gives for every $a \in \mathbb{Z}$ a unique quotient q and a unique remainder r. So we have maps

$$q_b \colon \mathbb{Z} \to \mathbb{Z},\ a \mapsto q \qquad \text{(the quotient by } b\text{)},$$
$$r_b \colon \mathbb{Z} \to \mathbb{N}_b,\ a \mapsto r \qquad \text{(the remainder after division by } b\text{)}.$$

Thus $a = q_b(a)b + r_b(a)$ for all $a \in \mathbb{Z}$. Theorem 8.1 states that the map $\mathbb{Z} \times \mathbb{N}_b \to \mathbb{Z}$, $(q, r) \mapsto qb + r$ is bijective. The inverse is $a \mapsto (q_b(a), r_b(a))$.

Algorithm

Here we consider only the case $a \in \mathbb{N}$. Instead of comparing multiples of b with a, one can subtract b repeatedly, starting from a. For division of a by b, start with the numbers 0 and a. Take the successor of the first and subtract b from the second (if $a \geq b$), repeat this with the resulting pair of numbers, and continue doing so until the second number is less than b. Then this second number is the remainder of a after division by b and the first is the quotient. We divide 253 by 41:

0	1	2	3	4	5	6
253	212	171	130	89	48	7

So the quotient equals 6 and the remainder equals 7.

Python

Given two natural numbers a and b we can determine whether $a \geq b$, and if so we can determine their difference. This we can use for division with remainder. For this we add to the module `integer.py` a function `quotres`:

integer.py

```
def quotres(a, b):
    q, d = 0, idiff(a, b)
    while d[2] == a: q, a, d = succ(q), d[0], idiff(d[0], b)
    return [q, a]
```

```
>>> quotres(234, 13)
[18, 0]
>>> quotres(1234, 13)
[94, 12]
```

8.2 The $\sum$-notation

Addition in $\mathbb{Z}$ is associative. Given an n-tuple of integers $a_0, \ldots, a_{n-1}$, i.e. a map $\mathbb{N}_n \to \mathbb{Z}$, we have the notion of *sum* of these integers since it is irrelevant how parentheses are placed, for example

$$\begin{aligned}(a_0 + a_1) + (a_2 + a_3) &= a_0 + (a_1 + (a_2 + a_3)) = a_0 + ((a_1 + a_2) + a_3) \\ &= (a_0 + (a_1 + a_2)) + a_3 = ((a_0 + a_1) + a_2) + a_2.\end{aligned}$$

So one may write $a_0 + a_1 + a_2 + a_3$ for short. More generally, the notation $a_0 + \cdots + a_{n-1}$ has a unique meaning. A more precise notation is $\sum_{k=0}^{n-1} a_k$, a notation which avoids the use of the $\cdots$. Even of a sequence of length 0 the sum can be taken: we just agree that this empty sum equals 0. Thus the meaning of $\sum_{k=0}^{n-1} a_k$ is given inductively:

$$\begin{cases} \displaystyle\sum_{k=0}^{-1} a_k = 0 \\ \displaystyle\sum_{k=0}^{n} a_k = \left(\sum_{k=0}^{n-1} a_k\right) + a_n. \end{cases}$$

This notation can be used more generally, namely in the case of a set with an associative and commutative operation $+$ for which there is a neutral element 0, so for an abelian monoid with the additive notation for its operation.

We can generalize this notation further to maps $I \to \mathbb{Z},\ i \mapsto a(i)$, where I is a finite set and not just a standard set $\mathbb{N}_n$:

$$\sum_{i \in I} a(i).$$

It can be defined as follows: choose a bijection $\mathbb{N}_n \to I,\ k \mapsto i_k$, and put:

$$\sum_{i \in I} a(i) = \sum_{k=0}^{n-1} a(i_k).$$

This is independent of the chosen bijection by commutativity of the addition. A set I used this way is called an *index set*.

For I the empty set this sum equals 0. This notation can be used more generally for abelian monoids with the additive notation for its operation.

8.2.1 Geometric progressions

8.4 Definition. A *geometric progression* (or *geometric sequence*) is a sequence of the form $a, ar, ar^2, \ldots$, where a and r are integers. Integers are the only numbers

we have so far. Later, when more numbers are available, these will be allowed as well. The number r is called the *common ratio* of the geometric progression.

> A geometric progression $a, ar, ar^2, \dots$ can be seen as the course of a under the transformation $x \mapsto xr$ of $\mathbb{Z}$.

Let $a, ar, ar^2, \dots$ be a geometric progression. Let s_n be the sum of the first n terms of the progression:

$$s_n = \sum_{k=0}^{n-1} ar^k = a + ar + \cdots + ar^{n-1}.$$

These sums form a sequence $s_0, s_1, s_2, \dots$ Apart from the obvious connection

$$s_{n+1} = s_n + ar^n$$

between s_{n+1} and s_n, we also have

$$s_{n+1} = \sum_{k=0}^{n} ar^k = a + \sum_{k=1}^{n} ar^k = a + rs_n.$$

So the sequence (s_n) is also determined by

$$\begin{cases} s_0 = 0, \\ s_{n+1} = rs_n + a \quad \text{for all } n \in \mathbb{N}, \end{cases}$$

that is: (s_n) is the course of 0 under the transformation $x \mapsto rx + a$. These observations lead to two theorems.

8.5 Theorem. *Let a and r be numbers (integers) with $r \neq 1$. Then the sum s_n of the first n terms of the geometric progression $a, ar, ar^2, \dots$ is given by*

$$s_n = a \cdot \frac{r^n - 1}{r - 1}.$$

Proof. We already saw that $s_n + ar^n = a + rs_n$. From this it follows that $(r-1)s_n = a(r^n - 1)$. □

> Here we only considered integers, because that is all we have. In general this holds in any integral domain. For $a = 1$ we have $(r-1)(r^{n-1} + \cdots + r^2 + r + 1) = r^n - 1$, that is $r^{n-1} + \cdots + r^2 + r + 1 = \frac{r^n - 1}{r-1}$. Multiplication of all terms by a, obviously results in a times their sum.

8.6 Theorem. *Let a and r be numbers (integers) with $r \neq 1$. Let the sequence (s_n) be defined by*

$$\begin{cases} s_0 = 0, \\ s_{n+1} = rs_n + a \quad \textit{for all } n \in \mathbb{N}, \end{cases}$$

Then for all $n \in \mathbb{N}$

$$s_n = a \cdot \frac{r^n - 1}{r - 1}.$$

PROOF. We already noted that the sequence (s_n) is the same as the sequence (s_n) from theorem 8.5. □

8.7 Example. In example 6.8 we saw that the sequence (a_n) determined by

$$\begin{cases} a_0 = 0 \\ a_{n+1} = 2a_n + 1 \quad \text{for all } n \in \mathbb{N} \end{cases}$$

coincides with the sequence given by $a_n = 2^n - 1$. This can be seen as a special case of theorem 8.6: $a = 1$ and $r = 2$.

8.2.2 Sums, subsets and partitions

The sum of as many numbers 1 as there are elements in A obviously is the number of elements of A:

$$\sum_{a \in A} 1 = \#(A).$$

Here A is the index set and the map from A to $\mathbb{Z}$ maps every element of A to 1.

For U a subset of a finite set A, $\chi_U : A \to \{0, 1\}$ the characteristic function of U on A, and f a function on A we have

$$\sum_{a \in A} \chi_U(a) f(a) = \sum_{a \in U} f(a)$$

and in particular

$$\sum_{a \in A} \chi_U(a) = \sum_{a \in U} 1 = \#(U).$$

If Φ is a partition of a finite set A, then every $a \in A$ is an element of exactly one $U \in \Phi$ and so for every $a \in A$:

$$\sum_{U \in \Phi} \chi_U(a) = 1.$$

For Φ a partition of a finite set A and f a function on A we have

$$\sum_{a \in A} f(a) = \sum_{U \in \Phi} \sum_{a \in U} f(a).$$

Here the sum is taken of the function values class by class and subsequently the sum is taken over all classes. Obviously we have in particular:

$$\#(A) = \sum_{U \in \Phi} \#(U).$$

We could have done this using characteristic functions as well:

$$\sum_{a\in A} f(a) = \sum_{a\in A}\sum_{U\in\Phi} \chi_U(a)f(a) = \sum_{U\in\Phi}\sum_{a\in A} \chi_U(a)f(a) = \sum_{U\in\Phi}\sum_{a\in U} f(a).$$

We changed the order of the sums. It can be seen as taking a sum over a product set in two ways. See below.

8.2.3 Double sums

Above we had a double sum and interchanged the sums. Why doesn't this have effect on the result? Suppose we have a function $f\colon \mathbb{N}_m \times \mathbb{N}_n \to \mathbb{Z}$. We take the sum of the function values over all pairs (i, j):

$$\sum_{(i,j)\in\mathbb{N}_m\times\mathbb{N}_n} f(i,j).$$

This sum has mn terms, see also exercise 8 of chapter 5 for a bijection $\mathbb{N}_{mn} \to \mathbb{N}_m \times \mathbb{N}_n$ (the inverse is $(i, j) \mapsto in + j$). So we take the sum of all terms

$$\begin{array}{ccccc} f(0,0) & f(1,0) & f(2,0) & \dots & f(m-1,0) \\ f(0,1) & f(1,1) & f(2,1) & \dots & f(m-1,1) \\ f(0,2) & f(1,2) & f(2,2) & \dots & f(m-1,2) \\ \vdots & \vdots & \vdots & & \vdots \\ f(0,n-1) & f(1,n-1) & f(2,n-1) & \dots & f(m-1,n-1) \end{array}$$

The sum of the numbers in the ith column (numbered from 0 to $n-1$) is

$$\sum_{j\in\mathbb{N}_n} f(i,j).$$

So the total sum is

$$\sum_{(i,j)\in\mathbb{N}_m\times\mathbb{N}_n} f(i,j) = \sum_{i\in\mathbb{N}_m}\sum_{j\in\mathbb{N}_n} f(i,j).$$

By first taking the sums of the rows we obtain

$$\sum_{(i,j)\in\mathbb{N}_m\times\mathbb{N}_n} f(i,j) = \sum_{j\in\mathbb{N}_n}\sum_{i\in\mathbb{N}_m} f(i,j).$$

So the sums can be interchanged:

$$\sum_{i\in\mathbb{N}_m}\sum_{j\in\mathbb{N}_n} f(i,j) = \sum_{j\in\mathbb{N}_n}\sum_{i\in\mathbb{N}_m} f(i,j).$$

8.3 The g-Adic Notation of Natural Numbers

So far we used the decimal notation of numbers. For the arithmetic we have not used this powerful notation, since every operation was constructed using the successor transformation only. Now we take a closer look at this notation. First we will see that it is based on division with remainder. Instead of taking the number ten we fix an arbitrary natural number g greater than 1 for the g-adic notation of natural numbers. The number g is called the *base* of this notation.

Let $a \in \mathbb{N}$. We divide by g, then we divide the quotient by g, and so on. Here to divide stands for to divide with remainder. After N divisions:

$$\begin{aligned}
a = a_0 &= a_1 g + c_0 && \text{with } a_1 \in \mathbb{N} \text{ and } c_0 \in \mathbb{N}_g,\\
a_1 &= a_2 g + c_1 && \text{with } a_2 \in \mathbb{N} \text{ and } c_1 \in \mathbb{N}_g,\\
a_2 &= a_3 g + c_2 && \text{with } a_3 \in \mathbb{N} \text{ and } c_2 \in \mathbb{N}_g,\\
&\vdots\\
a_{N-1} &= a_N g + c_{N-1} && \text{with } a_N \in \mathbb{N} \text{ and } c_{N-1} \in \mathbb{N}_g.
\end{aligned}$$

Thus we have

$$\begin{aligned}
a &= a_1 g + c_0 = a_2 g^2 + c_1 g + c_0 = a_3 g^3 + c_2 g^2 + c_1 g + c_0 = \cdots\\
&= a_{N-1} g^{N-1} + c_{N-2} g^{N-2} + \cdots + c_1 g + c_0. \qquad (8.1)
\end{aligned}$$

The sequence $a_0, a_1, a_2, \ldots$ satisfies

$$\begin{cases} a_0 = a, \\ a_{n+1} = q_g(a_n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

In other words the sequence $a_0, a_1, a_2, \ldots$ is the course of a under the transformation q_g of $\mathbb{N}$. For the iteration of a transformation see section 6.4.

8.8 Lemma. *For every $a \in \mathbb{N}$ there exists an $N \in \mathbb{N}$ such that $q_g^n(a) = 0$ for all $n \geq N$.*

Proof. Let $b \in \mathbb{N}$. Division by g gives $b = qg + r$ with $q \in \mathbb{N}$ and $r \in \mathbb{N}_g$. From

$$(b-q)g = bg - qg = bg - b + r = b(g-1) + r \geq b$$

it follows that $b > q$ if $b > 0$. So the course of an a under q_g is a sequence $a_0, a_1, a_2, \ldots$ with

$$a = a_0 > a_1 > a_2 > \cdots > a_{N-1} > 0 \quad \text{and } a_n = 0 \text{ for all } n \geq N. \qquad \square$$

8.9 Notation. Let $(x_0, x_1, \ldots, x_{n-1})$ be a sequence of natural numbers of length n. Then

$$[x_{n-1}, \ldots, x_1, x_0]_g = \sum_{i=0}^{n-1} x_i g^i = x_0 g^0 + x_1 g^1 + \cdots + x_{n-1} g^{n-1}.$$

This notation is also determined by

$$\begin{cases} [x_0]_g = x_0, \\ [x_1, x_0]_g = x_0 + x_1 g, \\ [x_n, x_{n-1}, \ldots, x_0]_g = [[x_n, x_{n-1}, \ldots, x_1]_g, x_0]_g \quad \text{for all } x_0, \ldots, x_n. \end{cases}$$

So for example $[5,3,2,4]_8 = [43,2,4]_8 = [346,4]_8 = [2772]_8 = 2772$. Notice the order: $[4,7,2]_{10} = 2 \cdot 10^0 + 7 \cdot 10^1 + 4 \cdot 10^2$. The notation $[4,7,2]_{10}$ becomes a notation for 472.

Using this notation the chain (8.1) of equalities can be written as follows:

$$a = [a_1, c_0]_g = [a_2, c_1, c_0]_g = [a_3, c_2, c_1, c_0]_g = \cdots = [a_{N-1}, c_{N-2}, \ldots, c_1, c_0]_g.$$

If $a_N = 0$, then $a_{N-1} < g$.

8.10 Definition. Let a be a natural number. If there exist natural numbers $a_0, a_1, \ldots, a_{n-1}$ with $a_i < g$ for $i = 0, 1, \ldots, n-1$ and, if $n \neq 1$, $a_{n-1} \neq 0$ such that

$$a = [a_{n-1}, \ldots, a_1, a_0]_g,$$

then we say that a can be written *g-adically*. The expression on the right hand side is called the *g-adic notation* of a.

We have seen by using repeatedly division with remainder that:

8.11 Proposition. *Every natural number can be written g-adically.* □

8.12 Notation. Let A be a set and $c \in A$. The set of sequences (a_n) in A is denoted by $\mathcal{R}(A)$. The subset of sequences (a_n) for which an $N \in \mathbb{N}$ exists with $a_n = c$ for all $n \geq N$ we denote by $\mathcal{R}_c(A)$. They are repeating sequences of the form

$$a_0, a_1, \ldots, a_{n-1}, \overline{c},$$

where $\overline{c}$ is the constant sequence $c, c, c, c, \ldots$. We say that such sequences have a *c-tail*.

In this section only the case $c = 0$ is used, but in, for example, decimal representations of rational numbers we will see other values of c. Think of $\frac{1}{3} = 0.333\cdots = 0.\overline{3}$.

We have a map

$$\mathcal{R}_0(\mathbb{N}) \to \mathbb{N},\ (a_n) \mapsto [a_{N-1}, \ldots, a_1, a_0]_g, \tag{8.2}$$

where for every sequence (a_n) an N is chosen such that $a_n = 0$ for all $n \geq N$. For $[a_{N-1}, \ldots, a_1, a_0]_g$ we will also write $[\ldots, a_2, a_1, a_0]_g$ since the choice of N is

irrelevant. The set $\mathcal{R}_0(\mathbb{N}_g)$ is the subset of $\mathcal{R}_0(\mathbb{N})$ consisting of sequences in $\mathbb{N}_g$ with a 0-tail. Restriction of the map (8.2) to this subset yields a map

$$\mathcal{R}_0(\mathbb{N}_g) \to \mathbb{N},\ (a_n) \mapsto [\ldots, a_2, a_1, a_0]_g. \tag{8.3}$$

Since every natural number can be written g-adically, the map (8.3) is surjective. We will see that it is also injective, in other words that the g-adic notation is unique. Clearly the map (8.2) is surjective: $(a, \overline{0}) \mapsto [a]_g = a$. It is not injective: $(g, \overline{0}) \mapsto [g]_g = g$ and $(0, 1, \overline{0}) \mapsto [1, 0]_g = g$.

We will show that the map (8.3) is bijective by giving an inverse. This inverse is the map

$$\mathbb{N} \to \mathcal{R}_0(\mathbb{N}_g),\ a \mapsto c_0, c_1, c_2, \ldots, \tag{8.4}$$

where $c_n = r_g(a_n)$ for all $n \in \mathbb{N}$, the sequence $a_0, a_1, a_2, \ldots$ being the course of a under the transformation q_g of $\mathbb{N}$. We have already seen that $a = [\ldots, c_2, c_1, c_0]_g$. It remains to show that, if $a = [\ldots, c_2, c_1, c_0]_g$, the numbers c_n are equal to $r_g(a_n)$. From $[\ldots, c_2, c_1, c_0]_g = [\ldots, c_2, c_1]_g \cdot g + c_0$ it follows that

$$q_g([\ldots, c_2, c_1, c_0]_g) = [\ldots, c_2, c_1]_g \quad \text{and} \quad r_g([\ldots, c_2, c_1, c_0]_g) = c_0.$$

So also $q_g^n([\ldots, c_2, c_1, c_0]_g) = [\ldots, c_{n+1}, c_n]_g$, that is $a_n = [\ldots, c_{n+1}, c_n]_g$, and indeed $r_g(a_n) = c_n$. We have shown:

8.13 Theorem. *Every natural number can be written g-adically and this notation is unique.* □

Algorithm

Determination of the g-adic notation of a natural number comes down to repeated division with remainder. The computation of the 8-adic notation of 851 by repeated division by 8:

$$\begin{aligned} 851 &= 106 \cdot 8 + 3 \\ 106 &= 13 \cdot 8 + 2 \\ 13 &= 1 \cdot 8 + 5 \\ 1 &= 0 \cdot 8 + 1 \end{aligned}$$

So $851 = [1, 5, 2, 3]_8$. The computation can be written as follows:

$$851 = [851]_8 = [106, 3]_8 = [13, 2, 3]_8 = [1, 5, 2, 3]_8.$$

In a scheme made from right to left:

0	1	13	106	851
	1	5	2	3

Left to a number in the top row the quotient after division by 8 and below the remainder after this division. The bottom row is the 8-adic notation.

Python

In Python the g-adic notation can be determined by division with remainder, collecting the remainders in a list. We use for this the data type `list`. The terms in such a list have an index, a natural number. Since we are representing natural numbers we will make no use of this and use only some simple operations on lists.

The function `repres(a,g)` returns the g-adic notation of a as a `list`:

integer.py

```
def repres(a, g):
    result = [a]
    while leq(g, result[0]): result[:1] = quotres(result[0], g)
    return result
```

```
>>> repres(2304, 10)
[2, 3, 0, 4]
>>> repres(2304, 16)
[9, 0, 0]
>>> repres(2304, 2)
[1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
>>> repres(2304, 100)
[23, 4]
```

And backwards:

integer.py

```
from functools import reduce

def nat(nrlist, g):
    def sumg(a, b): return isum(iprod(a, g), b)
    return reduce(sumg, nrlist)
```

```
>>> nat([2,0,3,6], 7)
713
>>> nat([2,0,3,6], 10)
2036
```

The 2-adic notation is also called the *binary* notation. Usually one writes words in 0 and 1: instead of $[1,1,0,1,0,1,0,0,1,0,0,1]_2$ one writes 110101001001 for short. It is understood that it is based on the binary notation.

Division by 8 gives only the numbers 0 to 7 as possible remainders. The 8-adic (= *octal*) notation of a number is a finite sequence of these numbers. The 8-adic notation of 3401 is $[6,5,1,1]_8$ or 6511 for short. In the octal notation numbers are words in the digits 0 to 7.

The 16-adic (= *hexadecimal*) notation of a number is a finite sequence consisting of numbers from 0 to 15:

$$3401 = [3401]_{16} = [212, 9]_{16} = [13, 4, 9]_{16}$$

For a compact hexadecimal notation also symbols for the numbers 10 to 15 are needed. Usually the letters A to F are used for this purpose. The hexadecimal notation of 3401 then is D49.

8.4 Arithmetic in a g-Adic System

The arithmetic (addition, subtraction, multiplication, division with remainder) in the decimal system is in a well-known way based on the arithmetic using only the symbols $0, 1, \ldots, 9$. More generally, the arithmetic in the g-adic system can be reduced to the arithmetic with the numbers $0, \ldots, g-1$. For doing calculations g-adically it is advantageous to know the multiplication tables of $0, \ldots, g-1$. For $g = 2$ only little remains to learn by heart; on the other hand many computations are needed: compare the addition of 268 and 341 in the decimal and the binary system:

268	100001100
341	101010101
609	1001100001

Algorithms

For addition, multiplication and division with remainder there are algorithms that are well-known in the decimal notation. For division with remainder there is the well-known method of long division. In these algorithms one computes with multiples of powers of the base number. It is all based on the rules of arithmetic for the natural numbers:

$$\begin{aligned}
268 + 341 &= [2, 6, 8]_{10} + [3, 4, 1]_{10} = 2 \cdot 10^2 + 6 \cdot 10 + 8 + 3 \cdot 10^2 + 4 \cdot 10 + 1 \\
&= (2 + 3) \cdot 10^2 + (6 + 4) \cdot 10 + (8 + 1) = [5, 10, 9]_{10} = [5 + 1, 0, 9]_{10} \\
&= [6, 0, 9]_{10} = 609. \\
48 \cdot 234 &= [4, 8]_{10} \cdot [2, 3, 4]_{10} = 4 \cdot [2, 3, 4, 0]_{10} + 8 \cdot [2, 3, 4]_{10} \\
&= [8, 12, 16, 0]_{10} + [16, 24, 32]_{10} = [9, 3, 6, 0]_{10} + [1, 8, 7, 2]_{10} \\
&= [10, 11, 13, 2]_{10} = [1, 2, 2, 3, 2]_{10}.
\end{aligned}$$

Python

First some simple conversions:

integer.py

```
def remove0(nrlist):
    while nrlist[0] == 0 and nrlist != [0]: del nrlist[0]
    return nrlist

def normalize(nrlist, g):
    normalform = []
    while nrlist != []:
        qr = quotres(nrlist.pop(), g)
        if nrlist != []: nrlist[-1] = isum(nrlist[-1], qr[0])
        elif qr[0] != 0: nrlist = [qr[0]]
        normalform.insert(0, qr[1])
    return remove0(normalform)

def equalize(nrlist1, nrlist2):
    list1, list2 = [0 for i in nrlist1], [0 for i in nrlist2]
    nrlist1 = list2 + nrlist1
    nrlist2 = list1 + nrlist2
    while nrlist1 != [0] and nrlist2 != [0] and\
    nrlist1[0] == nrlist2[0] == 0:
        del nrlist1[0]
        del nrlist2[0]
    return (nrlist1, nrlist2)
```

Their effect is shown in the examples:

```
>>> remove0([0, 0, 0, 5, 11, 3])
[5, 11, 3]
>>> normalize([0, 0, 0, 5, 11, 3], 8)
[6, 3, 3]
>>> equalize([3, 4, 34], [45, 0, 1, 1, 12])
([0, 0, 3, 4, 34], [45, 0, 1, 1, 12])
```

The sum and product of natural numbers:

integer.py

```
def gsum(nrlist1, nrlist2, g):
    (nrlist1, nrlist2) = equalize(nrlist1, nrlist2)
    return normalize([isum(nrlist1[i], nrlist2[i]) for i in\
    range(len(nrlist1))], g)

def gprod(nrlist1, nrlist2, g):
    products = [[iprod(i, j) for j in nrlist2] for i in nrlist1]
    def sum1(list1, list2): return gsum(list1 + [0], list2, g)
    return normalize(reduce(sum1, products), g)
```

Examples:

```
>>> gsum([3, 4, 34], [45, 0, 1, 1, 12], 8)
[5, 5, 0, 5, 2, 6]
>>> gprod([3, 4, 34], [45, 0, 1, 1, 12], 8)
[2, 6, 5, 3, 7, 2, 2, 5, 0]
```

We add to `integer.py` the algorithms for the ordering and the subtraction. The ordering:

integer.py

```
def lesseq(nrlist1, nrlist2):
    (nrlista, nrlistb) = equalize(nrlist1, nrlist2)
    while nrlista != [] and nrlista[0] == nrlistb[0]:
        del nrlista[0]
        del nrlistb[0]
    if nrlista == []: return True
    else: return leq(nrlista[0], nrlistb[0])
```

```
>>> lesseq([2, 0, 4, 4], [2, 0, 5, 1])
True
>>> lesseq([2, 0, 15, 4], [2, 0, 5, 1])
False
```

Subtraction:

integer.py

```
def sub0(nrlist1, nrlist2, g):
    (nrlist1, nrlist2) = equalize(nrlist1, nrlist2)
    diffs = []
    while nrlist1 != []:
        diffs.append(difference(isum(nrlist1.pop(0), g),\
        succ(nrlist2.pop(0))))
    diffs[-1] = succ(diffs[-1])
    diffs = normalize(diffs, g)
    diffs[:1] = []
    return diffs

def sub(nrlist1, nrlist2, g):
  return remove0(sub0(nrlist1, nrlist2, g))

def gmultiples(nrlist, g):
    mults = [[0, [0]]]
    h = difference(g, 1)
    while mults[-1][0] != h:
        s = succ(mults[-1][0])
```

```
>>> lesseq([2, 0, 4, 4], [2, 0, 5, 1])
True
>>> lesseq([2, 0, 15, 4], [2, 0, 5, 1])
False
>>> sub0([4, 2, 8, 0], [4, 2, 7, 11], 16)
[0, 0, 0, 5]
>>> sub([4, 2, 8, 0, 13], [4, 2, 7, 11], 16)
[3, 14, 5, 9, 2]
```

8.5 Direct Conversion Between Numeral Systems

A number given in the hexadecimal system can easily be converted to the decimal system since we are familiar with decimal arithmetic:

$$4\text{FE (hex)} = 4 \cdot 16^2 + 15 \cdot 16 + 14 = 1278.$$

If you are familiar with arithmetic in the hexadecimal system, you might prefer long divisions in that system.

A/4FE\7F	A/7F\C	A/C\1	A/1\0
46	78	A	0
9E	7	2	1
96			
8			

and so 4FE (hex) = 1278 (decimal). For doing hexadecimal arithmetic it is advisable to know the multiplication tables of 0 up to F, see Figure 8.1 on page 130. Especially for long divisions knowledge of these tables is advantageous.

Python

For direct conversion into another numeral system one performs the division algorithm in the given system repeatedly. For the conversion from g_1-adic to g_2-adic one does g_1-adic arithmetic, so g_2 has to be represented in the g_1-adic notation. In the division algorithm the multiple of the divisor has to be determined which will be subtracted from the dividend. For that purpose first a list of multiples of the divisor is made with `gmultiples`. The function `gdivmod0` is the step in the long division, which is repeatedly used. The long division is done by `gdivmod`.

integer.py

```
def gmultiples(nrlist, g):
    mults = [[0, [0]]]
    h = difference(g, 1)
    while mults[-1][0] != h:
        s = succ(mults[-1][0])
        mults.append([s, gprod([s], nrlist,g)])
    return mults

def gdivmod0(nrlist1, nrlist2, g):
    lista = [[0, [0]]]
    listb = gmultiples(nrlist2, g)[1:]
    while listb != [] and lesseq(listb[0][1], nrlist1):
        lista.append(listb.pop(0))
    return [lista[-1][0], sub0(nrlist1, lista[-1][1], g)]

def gdivmod(nrlist1, nrlist2, g):
    nrlist1copy = nrlist1[:]
    nrlist2copy = nrlist2[:]
    rlist = []
    dlist = []
    qlist = []
    while nrlist1copy != [] and nrlist2copy != []:
        rlist.append(nrlist1copy.pop(0))
        dlist.append(nrlist2copy.pop(0))
    if nrlist2copy != []: return [[0], nrlist1]
    else:
        while nrlist1copy != []:
            qr = gdivmod0(rlist, dlist, g)
            qlist.append(qr[0])
            rlist = qr[1]
            rlist.append(nrlist1copy.pop(0))
        qr = gdivmod0(rlist, dlist, g)
        qlist.append(qr[0])
        rlist = qr[1]
        return [remove0(qlist), remove0(rlist)]
```

```
>>> gmultiples([1, 4], 8)
[[0, [0]], [1, [1, 4]], [2, [3, 0]], [3, [4, 4]], [4, [6, 0]], [5, [7,
 4]], [6, [1, 1, 0]], [7, [1, 2, 4]]]
>>> gdivmod0([5, 1], [1, 4], 8)
[3, [0, 5]]
>>> gdivmod([5, 4, 6, 0, 1, 4], [1, 4], 8)
[[3, 5, 6, 5, 3], [1, 0]]
```

The conversion of one system into another consists of repeatedly applying the division algorithm.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2	0	2	4	6	8	A	C	E	10	12	14	16	18	1A	1C	1E
3	0	3	6	9	C	F	12	15	18	1B	1E	21	24	27	2A	2D
4	0	4	8	C	10	14	18	1C	20	24	28	2C	30	34	38	3C
5	0	5	A	F	14	19	1E	23	28	2D	32	37	3C	41	46	4B
6	0	6	C	12	18	1E	24	2A	30	36	3C	42	48	4E	54	5A
7	0	7	E	15	1C	23	2A	31	38	3F	46	4D	54	5B	62	69
8	0	8	10	18	20	28	30	38	40	48	50	58	60	68	70	78
9	0	9	12	1B	24	2D	36	3F	48	51	5A	63	6C	75	7E	87
A	0	A	14	1E	28	32	3C	46	50	5A	64	6E	78	82	8C	96
B	0	B	16	21	2C	37	42	4D	58	63	6E	79	84	8F	9A	A5
C	0	C	18	24	30	3C	48	54	60	6C	78	84	90	9C	A8	B4
D	0	D	1A	27	34	41	4E	5B	68	75	82	8F	9C	A9	B6	C3
E	0	E	1C	2A	38	46	54	62	70	7E	8C	9A	A8	B6	C4	D2
F	0	F	1E	2D	3C	4B	5A	69	78	87	96	A5	B4	C3	D2	E1

Figure 8.1: Hexadecimal multiplication table

integer.py

```
def convert(nrlist, g1, g2):
    result = [nrlist]
    g = repres(g2, g1)
    while lesseq(g, result[0]):
        result[:1] = gdivmod(result[0], g, g1)
    return [nat(i, g1) for i in result]
```

```
>>> convert([3, 5, 0, 5, 7], 8, 12)
[8, 7, 5, 3]
```

Addition and multiplication ultimately rest on taking the successor. No use is made of the decimal notation of natural numbers. Exploiting this notation provides fast algorithms, even so with the primitive means used here.

Exercises

1. Write the number 2008 (decimal notation) in the 7-adic notation. The remainder after division by ten is eight. Make this division as a long division in the 7-adic numeral system.

2. Multiply D6EE and A3A3 (hexadecimal notation). Do this in the hexadecimal system using the multiplication tables.

3. Let $[c_n, c_{n-1}, \ldots, c_0]_{10}$ be the decimal notation for an $a \in \mathbb{N}$. Prove:

   a and $c_0 + c_1 + \cdots + c_n$ have the same remainder after division by 9.

4. Have another look at exercise 2 of chapter 1. Now use theorem 8.6.

5. Let a and r be integers with $r \neq 1$. Let the sequence (t_n) be the course of a $c \in \mathbb{Z}$ under the transformation $x \mapsto rx + a$:

   $$\begin{cases} t_0 = c, \\ t_{n+1} = rt_n + a \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

   Determine a formula for t_n.

6. The sequence (a_n) is the course of 0 under the transformation $x \mapsto 7x + 3$ of $\mathbb{Z}$.
   (i) Determine the 7-adic notation of a_n.
   (ii) Which formula for a_n follows from theorem 8.6?

7. The sequence (a_n) is the course of 0 under the transformation $x \mapsto 7x + 11$ of $\mathbb{Z}$. Determine the 7-adic notation of a_n.

8. The 8-adic notation of an $a_n \in \mathbb{N}$ consists of $2n$ digits: $51515151 \cdots 51$, alternating 5 and 1.
   (i) What is the hexadecimal notation of a_n ?
   (ii) Which formula for a_n follows from theorem 8.5?

9. (i) Since 8 equals 2^3 there is a very simple connection between the octal and the binary notation. Conversion between these notations is easy. What is this connection?

   (ii) In the operating system Unix (and Linux) every file has an 'owner' and this owner is a member of a 'group'. The owner determines what others are allowed to do with his file. He does so for: the owner, the group and the world (= everybody). In each of these cases there are three actions which are permitted or not:

   `r` - reading of the file (then for example copying is possible),
   `w` - writing in the file (you may change it),
   `x` - the execution of the file (which makes sense for an executable file).

   To indicate whether something is off or on `0` and `1` will be used. Three of these bits are needed to indicate permissions. For example `110` means `rw-`: the file can be read and changed, but not executed. That is `6` $(= 4 + 2 + 0)$ in the octal system. For owner, group and world three of these digits are needed. Thus `640` means that the owner can read and write but not execute (convenient when it is not an executable, a document for example), members of the group can only read and others cannot do anything with the file.

   a) What is the meaning `764` here?

b) Describe with three octal digits that a computer program may be copied, changed and executed by the owner, that members of the group can copy and execute it, while by others it can only be executed.

The Unix command to set the permissions of a file `foo` to for example `640` is

```
chmod 640 foo
```

10. Eight bits form a byte. In the ASCII-code characters are given as bytes. Often a byte is written in the hexadecimal system. For a byte two hexadecimal digits are needed.

    (i) Bytes may be given in the decimal, binary, octal or hexadecimal system. What is `10011011` in these notations?

    True color means that colors displayed on the monitor are given in the RGB-code by three bytes, one for **R**ed, one for **G**reen and one for **B**lue. Sometimes the decimal and sometimes the hexadecimal notation is used. In the decimal notation there are three numbers from 0 up to 255, for example 210.35.106. In the hexadecimal notation this is `D2316A` (without the dots). In html-files for example `#D2316A` is used, with a `#` for the hexadecimal notation. Each of the three numbers stands for the intensity of the corresponding color.

    (ii) How many colors are there in the true color scheme?

    (iii) The RGB-code for white is `FFFFFF`. What are the RGB-codes for red, purple, black and gray?

    (iv) If the sum of two color codes equals `FFFFFF`, then the two colors are called *complementary* to each other. Which colors are complementary to red, green and blue?

11. Let the number a be given g-adically:

$$a = [c_n, \ldots, c_0]_g.$$

    (i) Show that $g^n \leq a < g^{n+1}$.

    (ii) Show that $c_n g^n \leq a < (c_n + 1)g^n$.

    (iii) How can the number c_n be determined without $c_0, \ldots, c_{n-1}$ being determined first?

    (iv) How can we determine the g-adic notation $[c_n, \ldots, c_0]_g$ of a from left to right, that is first c_n, then c_{n-1}, and so on?
    (The disadvantage of this method is that sufficiently many powers of g have to be computed first.)

12. **The Cantor representation**. Show that every natural number $\neq 0$ is uniquely representable as

$$c_1 \cdot 1! + c_2 \cdot 2! + c_3 \cdot 3! + \cdots + c_n \cdot n!,$$

    where for $i = 1, \ldots, n$ the number c_i is a natural number $\leq i$, and $c_n \neq 0$. (Hint: first determine n and c_n.)

13. Let $g \in \mathbb{N}$ with $g > 1$. We consider the directed graph having $\mathcal{R}_0(\mathbb{N})$ as vertex set

and as edges the ordered pairs

$$((a_0, a_1, \dots, a_k, a_{k+1}, a_{k+2}, \dots), (a_0, a_1, \dots, r_g(a_k), a_{k+1} + q_g(a_k), a_{k+2}, \dots)),$$

where $(a_0, a_1, a_2, \dots) \in \mathcal{R}_0(\mathbb{N})$ and $k \in \mathbb{N}$.

(i) Verify

$$[\dots, a_2, a_1, a_0]_g = [\dots, a_{k+2}, a_{k+1} + q_g(a_k), r_g(a_k), \dots, a_1, a_0]_g.$$

(ii) How many of the edges have a given vertex as tail?

(iii) Which vertices are both head and tail of a loop?

(iv) Starting at a vertex of the graph and walking along edges (in the direction of these edges) will you then ultimately arrive in a loop? (Hint: consider the sum of the terms of a sequence.)

(v) Walks starting at a given vertex, what do they have in common?

14. Prove that $\mathcal{R}_0(\mathbb{N})$ is countable.

15. Prove that $\mathcal{F}(\mathbb{N})$ is countable, see notation 6.6.

16. Let A be a finite set and B a countable set. Prove that B^A is countable. How can a bijection $\mathbb{N} \to B^A$ be easily made from bijections $f\colon \mathbb{N}_n \to A$ and $g\colon \mathbb{N} \to B$?

17. Prove that for all $n \in \mathbb{N}$

$$\sum_{k=0}^{n} (-1)^k \cdot k^2 = (-1)^n \cdot \sum_{k=0}^{n} k.$$

18. Let $a_0, a_1, a_2, \dots$ be a sequence of integers. The sequence of sums (s_n) of this sequence is given by $s_n = \sum_{k=0}^{n-1} a_k$. Let be given

$$s_1 = 1 \quad \text{and} \quad s_{100} = 100.$$

(i) Prove that there is an $m \in \mathbb{N}$ such that s_m is odd and s_{m+1} is even.

(ii) Prove that there is a $k \in \mathbb{N}$ such that a_k is odd and a_{k+1} is even.

9 The Rational Numbers

In chapter 7 the system $\mathbb{N}$ has been extended to the number system $\mathbb{Z}$ in order to make subtraction possible in all cases. The rules of arithmetic also hold in this extended number system. That in fact is the main reason for calling $\mathbb{Z}$ a number system. Now we are going to extend the number system further to make division by nonzero numbers possible. For extensions like these it is usually clear what we want to achieve, but it is not clear that such an extension actually exists. Maybe we want something impossible.

We constructed $\mathbb{Z}$ from $\mathbb{N}$ by seeing integers as differences of natural numbers. Similarly we will construct the rationals from $\mathbb{Z}$ by seeing them as quotients of integers, also called fractions. This construction is in fact applicable to any integral domain, but here we will confine to $\mathbb{Z}$.

The new property of the extended system, the system of rational numbers, is the possibility of division by nonzero elements. Such a structure is called a field. In this chapter we will have a first look at equations over a field in general.

The representation of a rational as a fraction of integers is not unique. The simplification of fractions leads in a natural way to the notion of greatest common divisor of integers. We will take a closer look at the simplification of fractions and the related computation of greatest common divisors.

In geometry numbers frequently occur as ratios of lengths of line segments. We will see that the system of the rationals is not sufficient for this purpose. More numbers are needed.

9.1 The Construction of $\mathbb{Q}$

In order to achieve that equations $m + x = n$ are solvable for all m and n we have extended $\mathbb{N}$ to $\mathbb{Z}$. Now we will extend $\mathbb{Z}$ further. We want all equations $bx = a$ (with $b \neq 0$) to be solvable and the rules of arithmetic to remain valid. Let's assume we have indeed such an extension of $\mathbb{Z}$. Let a_1, a_2, b_1, b_2 be integers with $b_1, b_2 \neq 0$ and let x_1 and x_2 be numbers in the extended number system such that $b_1 x_1 = a_1$ and $b_2 x_2 = a_2$. Since the rules of arithmetic are assumed to hold, we have

$$b_1 b_2 x_1 x_2 = a_1 a_2 \quad \text{and} \quad b_1 b_2 (x_1 + x_2) = a_1 b_2 + a_2 b_1,$$

so the number x_1x_2 is a solution of $b_1b_2x = a_1a_2$ and the number $x_1 + x_2$ of $b_1b_2x = a_1b_2 + a_2b_1$. Writing the solution of $bx = a$ as a 'fraction' $\frac{a}{b}$ we see that the collection of these fractions is all we need for addition and multiplication, the sum and the product of fractions being again such a fraction:

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{a_1b_2 + b_1a_2}{b_1b_2} \quad \text{and} \quad \frac{a_1}{b_1} \cdot \frac{a_2}{b_2} = \frac{a_1a_2}{b_1b_2}.$$

If $\mathbb{Q}$ is the desired extension of $\mathbb{Z}$, then we have a surjection

$$\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}) \to \mathbb{Q}, \quad (a, b) \mapsto \frac{a}{b}.$$

What then is the corresponding equivalence relation in $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$? If $b_1x = a_1$ as well as $b_2x = a_2$, then $b_1b_2x = a_1b_2$ and $b_1b_2x = a_2b_1$, and so $a_1b_2 = b_1a_2$. From all this it is clear how to construct the set $\mathbb{Q}$ and how to define addition and multiplication in this set.

9.1.1 The construction

Our starting point is the set of all pairs (a, b) of integers with $b \neq 0$.

9.1 Definition. For $(a_1, b_1), (a_2, b_2) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ we define

$$(a_1, b_1) \simeq (a_2, b_2) \iff a_1b_2 = b_1a_2.$$

This defines a relation $\simeq$ in $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.

9.2 Proposition. *The relation $\simeq$ is an equivalence relation.*

Proof. We prove that $\simeq$ is reflexive, symmetric and transitive.

reflexive: For $a \in \mathbb{Z}$ and $b \in \mathbb{Z} \setminus \{0\}$ we have $(a, b) \simeq (a, b)$, since this means $ab = ba$.

symmetric: From $(a_1, b_1) \simeq (a_2, b_2)$ follows $a_1b_2 = b_1a_2$ and this is equivalent to $(a_2, b_2) \simeq (a_1, b_1)$.

transitive: From $(a_1, b_1) \simeq (a_2, b_2)$ and $(a_2, b_2) \simeq (a_3, b_3)$ follows: $a_1b_2 = b_1a_2$ and $a_2b_3 = b_2a_3$. So

$$a_1b_2b_3 = b_1a_2b_3 = b_1b_2a_3$$

and since $b_2 \neq 0$ the cancellation law for the multiplication in $\mathbb{Z}$ implies $a_1b_3 = b_1a_3$, that is $(a_1, b_1) \simeq (a_3, b_3)$. □

9.3 Definition. $\mathbb{Q} = (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})/\simeq)$. We denote the class of $(a, b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by $\frac{a}{b}$. Elements of $\mathbb{Q}$ are called *rational* numbers. The expression $\frac{a}{b}$ is called a *fraction*. The a is the *numerator* of the pair (a, b) representing the fraction $\frac{a}{b}$ and b is its *denominator*.

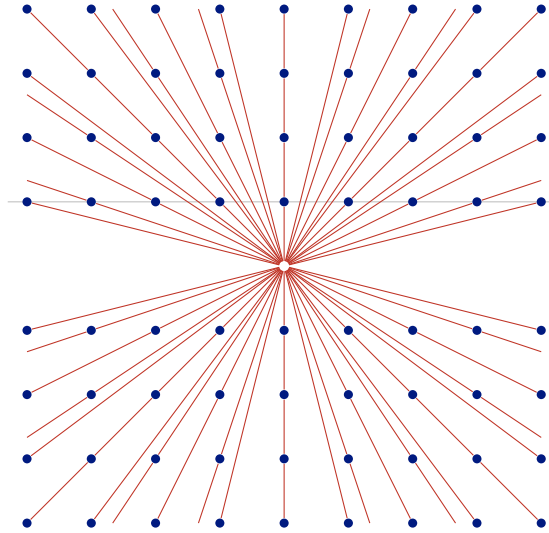


Figure 9.1: The partition $\mathbb{Q}$ of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$

Pairs (a_1, b_1) and (a_2, b_2) represent the same class if and only if $(a_1, b_1) \simeq (a_2, b_2)$. In other words the rational numbers $\frac{a_1}{b_1}$ and $\frac{a_2}{b_2}$ are equal if and only if $a_1 b_2 = a_2 b_1$.

Figure 9.1 is a picture of the partition $\mathbb{Q}$ of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$. From a geometrical point of view the classes are formed of the lattice points lying on the same straight line through the origin. Intersection of these lines with the line $y = 1$ defines a correspondence of the equivalence classes with points on this line. This gives a clear geometrical picture of the extension of $\mathbb{Z}$ to $\mathbb{Q}$.

In the introduction of this section it is shown how to add and multiply in the desired extension of $\mathbb{Z}$ *if such an extension exists*. So the definitions of addition and multiplication in definition 9.5 is not a surprise. The following lemma shows that these definitions are correct: they are independent of the choces of the representatives in the equivalence classes.

9.4 Lemma. *Let* (a_1, b_1), (a_2, b_2), (a_1', b_1'), $(a_2', b_2') \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ *be such that* $(a_1', b_1') \simeq (a_1, b_1)$ *and* $(a_2', b_2') \simeq (a_2, b_2)$. *Then*

$$(a_1' b_2' + b_1' a_2', b_1' b_2') \simeq (a_1 b_2 + b_1 a_2, b_1 b_2) \quad \textit{and} \quad (a_1' a_2', b_1' b_2') \simeq (a_1 a_2, b_1 b_2).$$

Proof. The integers $b_1 b_2$ and $b_1 b_2$ are nonzero. Since $a_1' b_1 = b_1' a_1$ and $a_2' b_2 = b_2' a_2$, we have

$$\begin{aligned}(a_1' b_2' + b_1' a_2') b_1 b_2 &= (a_1' b_2' b_1 b_2 + b_1' a_2') b_1 b_2 = a_1 b_2' b_1 b_2 + b_1' a_2 b_1 b_2 \\ &= a_1 b_2' b_1' b_2 + b_1' a_2 b_1 b_2' = (a_1 b_2 + b_1 a_2) b_1' b_2'\end{aligned}$$

and

$$a_1' a_2' b_1 b_2 = a_1 a_2 b_1' b_2'. \qquad \square$$

9.5 Definition. Addition in $\mathbb{Q}$:

$$\frac{a_1}{b_1} + \frac{a_2}{b_2} = \frac{a_1 b_2 + b_1 a_2}{b_1 b_2}.$$

Multiplication in $\mathbb{Q}$:

$$\frac{a_1}{b_1} \cdot \frac{a_2}{b_2} = \frac{a_1 a_2}{b_1 b_2}.$$

For the arithmetic with fractions the following rules are convenient.

9.6 Lemma. (i) $\dfrac{ac}{bc} = \dfrac{a}{b}$ *for* $a \in \mathbb{Z}$ *and* $b, c \in \mathbb{Z} \setminus \{0\}$.

(ii) $\dfrac{a}{b} + \dfrac{c}{b} = \dfrac{a+c}{b}$ *for* $a, c \in \mathbb{Z}$ *and* $b \in \mathbb{Z} \setminus \{0\}$.

Proof. (i) follows from $ac \cdot b = bc \cdot a$ and for (ii): $\dfrac{a}{b} + \dfrac{c}{b} = \dfrac{ab + bc}{b^2} = \dfrac{a+c}{b}$. $\square$

It still has to be shown that the rules of arithmetic hold in $\mathbb{Q}$.

9.7 Proposition. $\mathbb{Q}$ *is a commutative ring.*

Proof. The proof is straightforward. Because of its importance all axioms for a commutative ring will be checked. Let a, a_1, a_2 and a_3 be integers and b, b_1, b_2 and b_3 nonzero integers.

(G1) Associativity:

$$\begin{aligned}\left(\frac{a_1}{b_1} + \frac{a_2}{b_2}\right) + \frac{a_3}{b_3} &= \frac{a_1 b_2 + b_1 a_2}{b_1 b_2} + \frac{a_3}{b_3} = \frac{a_1 b_2 b_3 + b_1 a_2 b_3 + b_1 b_2 a_3}{b_1 b_2 b_3} \\ &= \frac{a_1}{b_1} + \frac{a_2 b_3 + b_2 a_3}{b_2 b_3} = \frac{a_1}{b_1} + \left(\frac{a_2}{b_2} + \frac{a_3}{b_3}\right).\end{aligned}$$

(G2) Zero element: $\dfrac{a}{b} + \dfrac{0}{1} = \dfrac{a+0}{b} = \dfrac{a}{b}$. So $\dfrac{0}{1}$ is the zero element.

(G3) Opposite: $\dfrac{a}{b} + \dfrac{-a}{b} = \dfrac{ab + b(-a)}{b} = \dfrac{0}{b^2} = \dfrac{0}{1}$.
So $\dfrac{-a}{b}$ is the opposite of $\dfrac{a}{b}$: $\quad -\dfrac{a}{b} = \dfrac{-a}{b}$.

(G4) Commutativity: $\dfrac{a_1}{b_1} + \dfrac{a_2}{b_2} = \dfrac{a_1 b_2 + b_1 a_2}{b_1 b_2} = \dfrac{a_2 b_1 + b_2 a_1}{b_2 b_1} = \dfrac{a_2}{b_2} + \dfrac{a_1}{b_1}$.

(R1) Associativity: $\left(\dfrac{a_1}{b_1} \cdot \dfrac{a_2}{b_2}\right) \cdot \dfrac{a_3}{b_3} = \dfrac{a_1 a_2 a_3}{b_1 b_2} = \dfrac{a_1}{b_1} \cdot \left(\dfrac{a_2}{b_2} \cdot \dfrac{a_3}{b_3}\right)$.

(R2) Unity element: $\frac{a}{b} \cdot \frac{1}{1} = \frac{a}{b}$. So $\frac{1}{1}$ is the unity element.

(R3) Distributivity:

$$\frac{a_1}{b_1} \cdot \left(\frac{a_2}{b_2} + \frac{a_3}{b_3}\right) = \frac{a_1}{b_1} \cdot \frac{a_2 b_3 + b_2 a_3}{b_2 b_3} = \frac{a_1 a_2 b_3 + a_1 b_2 a_3}{b_1 b_2 b_3}$$
$$= \frac{a_1 a_2 b_3}{b_1 b_2 b_3} + \frac{a_1 b_2 a_3}{b_1 b_2 b_3} = \frac{a_1 a_2}{b_1 b_2} + \frac{a_1 a_3}{b_1 b_3} = \frac{a_1}{b_1} \cdot \frac{a_2}{b_2} + \frac{a_1}{b_1} \cdot \frac{a_3}{b_3}.$$

(R4) Commutativity: $\frac{a_1}{b_1} \cdot \frac{a_2}{b_2} = \frac{a_1 a_2}{b_1 b_2} = \frac{a_2}{b_2} \cdot \frac{a_1}{b_1}$. □

9.1.2 $\mathbb{Z}$ as part of $\mathbb{Q}$

After $\mathbb{Z}$ was constructed we considered the elements $[n, 0]$ with $n \in \mathbb{N}$. Together they form a number system that can replace the $\mathbb{N}$ we started with. Here we have a similar situation: $\mathbb{Z}$ can be replaced by the subset of all $\frac{a}{1}$. This follows from:

$$\frac{a}{1} = \frac{b}{1} \iff a = b \quad (\text{for all } a, b \in \mathbb{Z}),$$

$$\frac{a}{1} + \frac{b}{1} = \frac{a+b}{1} \quad (\text{for all } a, b \in \mathbb{Z})$$

and

$$\frac{a}{1} \cdot \frac{b}{1} = \frac{ab}{1} \quad (\text{for all } a, b \in \mathbb{Z}).$$

Instead of $\frac{a}{1}$ we will simply write a.

9.1.3 $\mathbb{Q}$ is a field

The number system $\mathbb{Q}$ is a commutative ring with a special property:

$$\text{for every } r \in \mathbb{Q} \text{ with } r \neq 0 \text{ there is an } s \in \mathbb{Q} \text{ such that } rs = 1.$$

If $\frac{a}{b} = 0 \ (= \frac{0}{1})$, then $a = a \cdot 1 = b \cdot 0 = 0$ and vice versa, if $a = 0$, then $\frac{a}{b} = 0$. So if $r = \frac{a}{b} \neq 0$, then $a \neq 0$ and for $s = \frac{b}{a}$ we have:

$$rs = \frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab} = \frac{1}{1} = 1.$$

9.8 Definition. An element r of a ring R is called *invertible* if there is an $s \in R$ such that $rs = sr = 1$. The element s is called the *inverse* of r and is denoted by r^{-1}.

The invertible elements of a ring R form a group under multiplication. The inverse in the ring is the same as the inverse in this group, see definition 7.38.

9.9 Notation. The group of invertible elements of a ring R we denote by R^*.

In the ring $\mathbb{Z}$ only 1 and -1 have inverses. So $\mathbb{Z}^* = \{1, -1\}$. We have seen that in $\mathbb{Q}$ all nonzero elements have inverses: $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$.

9.10 Definition. A *field* is a commutative ring with $1 \neq 0$ and in which every element $\neq 0$ has an inverse.

If K is a field, then $K \setminus \{0\} = K^*$ and this is an abelian group. The group operation is the multiplication.

Instead of rs^{-1} we also write $\frac{r}{s}$. The fraction bar stands for 'divided by', which is in accordance with the identification of $\frac{a}{1}$ with the integer a. We have seen that in $\mathbb{Q}$ nonzero elements have inverses and so:

9.11 Theorem. *$\mathbb{Q}$ is a field.* □

In mathematics fields are important. In the chapters 17 and 19 the fields $\mathbb{R}$ (of the real numbers) and $\mathbb{C}$ (of the complex numbers) will be constructed. There are also finite fields. In chapter 13 we will see examples of finite fields. For fields often the letter K is used. That goes back to the German word Körper which means body or corps and in many languages the word for field is a direct translation of this German word, but not so in English and neither in Russian.

Note that we achieved equations $sx = r$ with $s \neq 0$ to be solvable. The unique solution is $x = s^{-1}r$. If $r = 0$, then the unique solution is $x = 0$, that is a field has no zero divisors. So fields are integral domains, but integral domains need not be fields: $\mathbb{Z}$ is an integral domain, but is not a field.

9.1.4 Exponentiation in $\mathbb{Q}$

For every $r \in \mathbb{Q}$ with $r \neq 0$ there is an inverse of r. We denote it by r^{-1}. Besides exponents in $\mathbb{N}$ we now also have -1 as an exponent. We will extend this to exponents in $\mathbb{Z}$. First a lemma.

9.12 Lemma. *Let $r \in \mathbb{Q}^*$ and $n \in \mathbb{N}$. Then $(r^{-1})^n = (r^n)^{-1}$.*

Proof. To prove that $(r^{-1})^n$ is the inverse of r^n. This follows from:

$$(r^{-1})^n \cdot r^n = (r^{-1}r)^n = 1^n = 1.$$ □

Integers are introduced as differences of natural numbers.

9.13 Lemma. *For $r \in \mathbb{Q}^*$ and $a \in \mathbb{Z}$ we have: if $a = n - m = q - p$ with $m, n, p, q \in \mathbb{N}$, then*

$$r^n(r^{-1})^m = r^q(r^{-1})^p.$$

Proof. From $n+p=q+m$ it follows that $r^n r^p = r^q r^m$ and so $r^n (r^m)^{-1} = r^q (r^{-1})^p$. The lemma now follows using lemma 9.12. □

As a consequence we can define r^a for integers a as follows.

9.14 Definition. Let $r \in \mathbb{Q}^*$ and $a \in \mathbb{Z}$ with $a = n - m$, where $m, n \in \mathbb{N}$. Then we define

$$r^a = r^n (r^{-1})^m.$$

In particular r^a has for $a \in \mathbb{N}$ the same meaning as before. Now we also have for $a \in \mathbb{N}$ the following meaning of r^{-a}: it is the a-th power of r^{-1}. We could have used this as a definition, but the more general approach has the advantage that the rules of arithmetic are more easily verified.

9.15 Proposition. *Let r and s be rational numbers $\neq 0$ and let a and b be integers. Then*

(i) $r^a r^b = r^{a+b}$,
(ii) $(r^a)^b = r^{ab}$,
(iii) $(rs)^a = r^a s^a$.

Proof. We write $a = n - m$ and $b = q - p$ with $m, n, p, q \in \mathbb{N}$.

(i) $r^a r^b = r^{n-m} r^{q-p} = r^n (r^{-1})^m r^q (r^{-1})^p = r^{n+q} r^{-1})^{m+p} = r^{a+b}$.
(ii) $(r^a)^b = (r^n (r^{-1})^m)^q (r^m (r^{-1})^n)^p = r^{nq} (r^{-1})^{mq} r^{mp} (r^{-1})^{np}$
$= r^{nq+mp} (r^{-1})^{mq+np} = r^{ab}$.
(iii) $(rs)^a = (rs)^n ((rs)^{-1})^m = r^n s^n ((s^{-1}) r^{-1})^m = r^n (r^{-1})^m s^n (s^{-1})^m = r^a s^a$.

□

9.1.5 The ordering of $\mathbb{Q}$

We see $\mathbb{Q}$ as an extension of $\mathbb{Z}$. The ordering $\leq$ of $\mathbb{Z}$ can be extended to $\mathbb{Q}$.

9.16 Definition. Let r and s be rational numbers. We can write r and s as fractions having a common denominator:

$$r = \frac{a}{b} \text{ and } s = \frac{c}{b} \text{ with } b > 0.$$

We define

$$r \leq s \iff a \leq c.$$

The relation $\leq$ in $\mathbb{Q}$ is well defined: if also $r = \frac{a'}{b'}$ and $s = \frac{c'}{b'}$ with $b' > 0$, then $a'b = ab'$ and $c'b = cb'$ and we have

$$a \leq c \iff ab' \leq cb' \iff a'b \leq c'b \iff a' \leq c'.$$

9.17 Proposition. *The relation $\leq$ is an ordering of $\mathbb{Q}$.*

Proof. Write the rational numbers under consideration as fractions having common denominators. Then the proof is easy. □

> On $\mathbb{Q}$ we also have an absolute value. This is important for the notion of limit which is crucial for the further extension of the field of rationals. Limits are studied in chapter 16.

Rational numbers always lie between integers, in fact:

9.18 Lemma. *Let $x \in \mathbb{Q}$. Then there is a unique $m \in \mathbb{Z}$ such that $m \leq x < m+1$.*

Proof. Let $x = \frac{a}{b}$ with $b \in \mathbb{N}^+$. Division with remainder gives $a = qb + r$ with $q \in \mathbb{Z}$ and $r \in \mathbb{N}_b$. Then $qb \leq a < (q+1)b$, that is $q \leq x < q+1$. So take $m = q$. The uniqueness follows from the uniqueness of the quotient after division with remainder. □

9.19 Definition. Let $x \in \mathbb{Q}$. Then the unique $m \in \mathbb{Z}$ such that $m \leq x < m+1$ is called the *floor* of x (or also the *entier* of x). Notation: $\lfloor x \rfloor = m$. (Also the notation $[x]$ is widely used.) The unique $n \in \mathbb{Z}$ with $n - 1 < x \leq n$ is called the *ceiling* of x. Notation: $\lceil x \rceil = n$.

> Thus for $x \in \mathbb{Z}$ we have $\lfloor x \rfloor = x = \lceil x \rceil$. For nonintegral rationals x we have $\lfloor x \rfloor < x < \lfloor x \rfloor + 1 = \lceil x \rceil$. Division with remainder was used in the definition of the floor of a rational number. Conversely we have: $q_b(a) = \lfloor \frac{a}{b} \rfloor$ and $r_b(a) = a - \lfloor \frac{a}{b} \rfloor b$.

9.2 Equations

9.20 Terminology. Equations of the form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = 0,$$

where $a_0, \ldots, a_n$ are elements of a field (like $\mathbb{Q}$) are called *polynomial equations*. The expression on the left hand side of the equation is called a *polynomial* in x over the field. If $a_n \neq 0$, then the polynomial and the equation are said to be of *degree* n. The coefficient a_n is called the *leading coefficient* of the polynomial. Equations of degree 1 are called *linear*, equations of degree 2 are called *quadratic* and equations of degree 3 *cubic*.

9.2.1 Linear equations

Linear equations over a field have a unique solution in that field. The equation

$$ax + b = 0$$

with $a \neq 0$ has the unique solution $x = -\frac{b}{a}$: if $ax + b = 0$, then $ax = -b$ and so $x = -\frac{b}{a}$. If a and b are rational numbers, then so is $-\frac{b}{a}$.

Abu Ja'far Muhammad ibn Musa Al-Khwarizmi (Baghdad? ±780 – Baghdad? ±850)

Al-Khwarizmi was the first to describe the solution of quadratic equations in a systematic algebraic manner. He did so in his book *Hisab al-jabr w'al-muqabala*. He probably knew Euclid's geometrical completion of the square. The word 'algebra' comes from 'al-jabr' and our word 'algorithm' is derived from 'Al-Khwarizmi'.

9.2.2 Quadratic equations

The solution of a quadratic equation is given by the well-known quadratic formula. For the solution of a quadratic equation there is a general recipe and when applied to the general equation $ax^2 + bx + c = 0$ with $a \neq 0$ the quadratic formula is the result. The recipe is known as the method of *completing the square* and applied to the general quadratic equation over the field $\mathbb{Q}$ it goes as follows:

1. Divide by a:
$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

2. Introduce a new unknown: $y = x + \frac{b}{2a}$. Equivalently: substitute $y - \frac{b}{2a}$ for x. Then:
$$\left(y - \frac{b}{2a}\right)^2 + \frac{b}{a}\left(y - \frac{b}{2a}\right) + \frac{c}{a} = 0,$$
that is
$$y^2 - \frac{b}{a}y + \frac{b^2}{4a^2} + \frac{b}{a}y - \frac{b^2}{2a^2} + \frac{c}{a} = 0.$$
After simplification:
$$y^2 = \frac{b^2}{4a^2} - \frac{c}{a} = \frac{b^2 - 4ac}{4a^2}.$$

3. If $\frac{b^2-4ac}{4a^2}$ is a square (in $\mathbb{Q}$), or, what amounts to the same, if $b^2 - 4ac$ is a square, then the equation is solvable in $\mathbb{Q}$ and otherwise it is not.

4. Suppose $b^2 - 4ac$ is a square, say $b^2 - 4ac = d^2$. Then the equation becomes $y^2 = (\frac{d}{2a})^2$, that is $y^2 - (\frac{d}{2a})^2 = 0$. The left hand side can be written as a product:
$$\left(y - \frac{d}{2a}\right)\left(y + \frac{d}{2a}\right) = 0.$$
Since a field has no zero divisors, it follows that $y - \frac{d}{2a} = 0$ or $y + \frac{d}{2a} = 0$, that is $y = \frac{d}{2a}$ or $y = -\frac{d}{2a}$.
5. Since $x = -\frac{b}{2a} + y$, the solutions are $x = -\frac{b}{2a} + \frac{d}{2a}$ and $x = -\frac{b}{2a} - \frac{d}{2a}$:
$$x = \frac{-b \pm d}{2a}.$$

Since $d^2 = b^2 - 4ac$, this can be written as:
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$
the formula for the solution of the quadratic equation $ax^2 + bx + c = 0$.

Remarks

a) The number of solutions depends on the number $b^2 - 4ac$. If it is 0, there is exactly one solution. If $b^2 - 4ac$ is a square $\neq 0$, then there are two solutions and if it is not a square, then there are no solutions. In chapter 10 we will see which rational numbers actually are squares. The number $b^2 - 4ac$ is called the *discriminant* of the equation.

b) While completing the square, $\frac{b}{a}$ was divided by 2. What is 2 in an arbitrary field? Simply, $2 = 1 + 1$. But division is only possible if $2 \neq 0$. There are fields in which $2 = 0$, the so-called fields of characteristic 2. In such fields one has $2a = 0a = 0$, that is $a + a = 0$, or $a = -a$. In chapter 13 we will see an example: a field with only two elements. For quadratic equations over fields of characteristic 2 the recipe of completing the square is not applicable and the quadratic formula is useless since it has 0 in the denominator of the fraction.

c) In chapter 17 the field $\mathbb{R}$ of the real numbers will be constructed. Real numbers are the kind of numbers many people experience as being 'real'. We will see the well-known fact that the nonzero squares in $\mathbb{R}$ are just the positive numbers. That is why one usually looks at the sign of d for the number of solutions.

d) Also in chapter 18 the field $\mathbb{Q}$ will be extended, but there in a very unusual way. Fields $\mathbb{Q}_p$ of p-adic numbers will be constructed, one for each prime number p. We will determine the squares in these fields as well.

e) In chapter 19 we extend $\mathbb{R}$ further to the field $\mathbb{C}$ of complex numbers. In that field all numbers are squares and therefore each quadratic equation is solvable in that field.

f) In chapter 13 we will consider finite fields consisting of p elements, one for each prime number p. In chapter 14 we investigate which of the elements in these fields are squares.

9.2.3 Number of solutions

The following theorem is about the number of solutions of an equation of degree n over a field. It is frequently used.

9.21 Theorem. *Let $n \in \mathbb{N}$. An equation of degree n over a field has at most n solutions.*

Proof. We will prove the theorem by mathematical induction on n. For $n = 0$ it is trivially true (and for $n = 1$ there is exactly 1 solution).

Suppose that for some $n \in \mathbb{N}$ equations of degree n have at most n solutions. Then to prove that equations of degree $n + 1$ have at most $n + 1$ solutions. We consider an equation of degree $n + 1$:

$$a_{n+1}x^{n+1} + a_n x^n + \cdots + a_2 x^2 + a_1 x + a_0 = 0, \tag{9.1}$$

where $a_0, \ldots, a_{n+1}$ are elements of a field ($\mathbb{Q}$ for example) and $a_{n+1} \neq 0$. If this equation has no solutions, then we are finished. If $x = a$ is a solution, that is

$$a_{n+1}a^{n+1} + a_n a^n + \cdots + a_2 a^2 + a_1 a + a_0 = 0,$$

then after subtraction

$$a_{n+1}(x^{n+1} - a^{n+1}) + a_n(x^n - a^n) + \cdots + a_2(x^2 - a^2) + a_1(x - a) = 0.$$

For every $k \in \mathbb{N}^+$ we have

$$x^k - a^k = (x - a) \cdot \sum_{i=0}^{k-1} x^i a^{k-1-i},$$

as is easily seen by elaborating the right hand side. It is also a consequence of theorem 8.5: for $a \neq 0$ it is equivalent with

$$r^k - 1 = (r - 1) \cdot \sum_{i=0}^{k-1} r^i,$$

where $r = \frac{x}{a}$. Writing $g_k(x) = \sum_{i=0}^{k-1} x^i a^{k-1-i}$ the equation becomes

$$(x - a)(a_{n+1}g_{n+1}(x) + a_n g_n(x) + \cdots + a_1 g_1(x)) = 0. \tag{9.2}$$

The equation

$$a_{n+1}g_{n+1}(x) + a_n g_n(x) + \cdots + a_1 g_1(x) = 0 \tag{9.3}$$

is of degree n and, therefore, has at most n solutions. Because a field has no zero divisors, any solution of equation (9.2) is a solution of equation (9.3) or of $x - a = 0$. So equation (9.2), which is equivalent to equation (9.1), has at most $n + 1$ solutions.

By mathematical induction it follows that for every n an equation of degree n has at most n solutions. □

If $x = p$ is a solution of the quadratic equation $ax^2 + bx + c = 0$, then by the above method the equation can be rewritten as:

$$ax^2+bx+c = a(x^2-p^2)+b(x-p) = (x-p)(a(x+p)+b) = a(x-p)(x+p+\tfrac{b}{a}) = 0.$$

So $x = -p - \frac{b}{a}$ is a solution and there are not more solutions. In chapter 19 the number system will be extended to $\mathbb{C}$, the field of complex numbers. Over $\mathbb{C}$ all polynomial equations of degree ≥ 1 have a solution. This is known as the *Fundamental Theorem of Algebra*. A proof is given in subsection 19.4.3.

9.3 Simplifying Fractions

The representation of a rational number as a fraction is not unique. Instead of $\frac{2}{3}$ you may also write $\frac{2002}{3003}$, being a more complicated expression. This section is about the opposite: simplifying fractions.

9.3.1 Divisors

9.22 Definition. Let a and d be integers. The integer d is called a *divisor* of a if there is an integer x such that $dx = a$. Notation: $d \mid a$. We also say: a is a *multiple* of d. If d is not a divisor of a, then we denote this as: $d \nmid a$.

$$\begin{array}{ll} 2 \mid 6, & \text{since } 2 \cdot 3 = 6. \\ 2 \mid -6, & \text{since } 2 \cdot (-3) = -6. \\ -2 \mid 8, & \text{since } (-2) \cdot (-4) = 8. \\ 12 \mid 0, & \text{since } 12 \cdot 0 = 0. \\ 1 \mid -17, & \text{since } 1 \cdot (-17) = -17. \\ 0 \mid 0, & \text{since } 0 \cdot 157 = 0. \end{array}$$

Note that '$2 \mid a$' means the same as 'a is even', and that '$2 \nmid a$' means that a is odd.

For a and b integers, the expressions '$b \mid a$' and '$\frac{a}{b}$' have a different meaning: the first is a proposition concerning numbers a and b and the second is not, it stands for a number. If $b \neq 0$, we have

$$b \mid a \iff \tfrac{a}{b} \in \mathbb{Z}.$$

We will derive some simple rules for divisors. The straightforward proofs demonstrate the use of the relation $\mid$.

9.23 Lemma. *Let a be an integer. Then:*

$$\text{(i) } 1 \mid a, \quad \text{(ii) } a \mid a, \quad \text{(iii) } a \mid 0.$$

Proof.

(i) $1 \cdot a = a$. So there is an integer x with $1 \cdot x = a$, that is $1 \mid a$.
(ii) $a \cdot 1 = a$. So $a \mid a$.
(iii) $a \cdot 0 = 0$. So $a \mid 0$. □

9.24 Proposition. *Let a, b and c be integers. Then:*

(i) *if $a \mid b$ and $b \mid c$, then $a \mid c$;*
(ii) *if $a \mid b$ and $a \mid c$, then $a \mid b + c$;*
(iii) *if $ca \mid cb$ and $c \neq 0$, then $a \mid b$.*

Proof.

(i) Suppose $a \mid b$ and $b \mid c$. Then there are integers x and y such that $b = ax$ and $c = by$. Then $c = axy$ and so $a \mid c$, since the integer $z = xy$ satisfies $c = az$.
(ii) Suppose $a \mid b$ and $a \mid c$. Then there are integers x, y such that $b = ax$ and $c = ay$. Then $b + c = ax + ay = a(x + y)$ and so $a \mid b + c$.
(iii) Suppose $ca \mid cb$ and $c \neq 0$. There is an integer x such that $cax = cb$, or $c(ax - b) = 0$. Since $c \neq 0$ it follows that $ax - b = 0$, or $ax = b$. So $a \mid b$. □

9.25 Proposition. *Let m and n be natural numbers with $m \mid n$ and $n \mid m$. Then $m = n$.*

Proof. We distinguish two cases.

a) $m = 0$. Then also $n = 0$, because $m \mid n$. And so $m = n$.
b) $m \neq 0$. Then also $n \neq 0$, because $n \mid m$. There are natural numbers x, y with $n = mx$ and $m = ny$. It follows that $n = mx = nyx$ and since $n \neq 0$ we have $yx = 1$. So $x = y = 1$, and therefore $m = n$. □

9.3.2 The greatest common divisor

A nonzero integer a has only finitely many divisors. The number 0 has infinitely many: every integer is a divisor of 0. On the other hand an integer $a \neq 0$ has infinitely many multiples, whereas only 0 is a multiple of 0. In this section we consider the divisors which integers have in common.

9.26 Definition. Let a and b be integers. We say that $d \in \mathbb{Z}$ is a *common divisor* of a and b if

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

If $a \neq 0$ or $b \neq 0$ then there are only finitely many common divisors of a and b. Then there also is a maximal one. This one we call the *greatest common divisor* of a and b, and denote it by $\gcd(a, b)$. If $\gcd(a, b) = 1$, then a and b are said to be *relatively prime*.

9.27 Example. The divisors of 24 are: ±1, ±2, ±3, ±4, ±6, ±8, ±12, ±24. The divisors of -18 are: ±1, ±2, ±3, ±6, ±9, ±18. So the common divisors of 24 and -18 are: ±1, ±2, ±3, ±6. Hence 6 is the greatest common divisor.

Since

$$\gcd(a,b) = \gcd(\pm a, \pm b),$$

it suffices to consider greatest common divisors of natural numbers. It also suffices to consider positive common divisors only when determining the greatest.

9.3.3 Application to fractions

Given a fraction $\frac{a}{n}$, both the numerator and the denominator may be divided by $\gcd(a,n)$, say $\frac{a}{\gcd(a,n)} = a'$ and $\frac{n}{\gcd(a,n)} = n'$. Then

$$\frac{a}{n} = \frac{a'}{n'}.$$

And then $\gcd(a',n') = 1$: if $d \mid a'$ and $d \mid n'$, then $d{\cdot}\gcd(a,n) \mid a$ and $d{\cdot}\gcd(a,n) \mid n$ and so $d \cdot \gcd(a,n) \leq \gcd(a,n)$, that is $d \leq 1$.

9.28 Definition. Let r be a rational number. Let $n \in \mathbb{N}^+$ satisfy $nr \in \mathbb{Z}$. Then n is said to be a *denominator* of r.

So we have: if $n \in \mathbb{N}^+$ is a denominator of r, then $r = \frac{nr}{n}$ with nr an integer. There exists a denominator n' such that $\gcd(n'r, n') = 1$. We will see that this denominator is in fact the least.

9.4 The Euclidean Algorithm

Clearly, in principle the greatest common divisor can be calculated: determine all divisors of the two numbers and take the greatest they have in common. That can be a lot of work and for large numbers it might be practically impossible. A fast way to determine the greatest common divisor was already known in ancient Greece. First a lemma.

9.29 Lemma. *Let a and b be integers, not both 0, and let $t \in \mathbb{Z}$. Then*

$$\gcd(a + tb, b) = \gcd(a, b).$$

Proof. A common divisor of a and b is a also common divisor of $a + tb$ and b, and vice versa. So the common divisors of a and b coincide with those of $a + tb$ and b. In particular in both cases the greatest common divisors are the same. □

9.30 Algorithm (Euclid). By using division with remainder, the calculation of the greatest common divisor of $a, b \in \mathbb{N}$ can, if $b \neq 0$, be reduced by lemma 9.29 to the calculation of the greatest common divisor of b and $a - qb$, where $0 < a - qb < b$. This reduction can be repeated as long as the second number is not 0. If the second number is 0, the first one is the greatest common divisor. This algorithm has been described by Euclid and it is known as the *Euclidean algorithm.*

In detail, for given natural numbers a and b, not both zero, the algorithm is as follows. If $b \neq 0$, then apply division with remainder:

$$a = q_1 b + r_1 \quad \text{with} \quad 0 \leq r_1 < b.$$

Thus the pair (a, b) is replaced by the pair (b, r_1) and $r_1 < b$. If $r_1 \neq 0$, then again by division with remainder

$$b = q_2 r_1 + r_2 \quad \text{with} \quad 0 \leq r_2 < r_1$$

and if $r_2 \neq 0$, then again by division with remainder

$$r_1 = q_3 r_2 + r_3 \quad \text{with} \quad 0 \leq r_3 < r_2.$$

As long as the remainder is nonzero, division with remainder is applied. The remainders form a strict descending sequence of natural numbers: if a remainder is $\neq 0$, then the remainder obtained in the next step is less. This process ends as soon as a remainder 0 occurs:

$$\begin{aligned} r_{n-2} &= q_n r_{n-1} + r_n \quad \text{with} \quad 0 < r_n < r_{n-1}, \\ r_{n-1} &= q_{n+1} r_n + 0. \end{aligned}$$

Then the greatest common divisor of a and b equals r_n (= the last nonzero remainder). This is a very fast algorithm. It does not involve the determination of divisors in advance.

9.31 Example. We compute the greatest common divisor of 1665 and 978:

$$\begin{aligned} 1665 &= 1 \cdot 978 + 687 \\ 978 &= 1 \cdot 687 + 291 \\ 687 &= 2 \cdot 291 + 105 \\ 291 &= 2 \cdot 105 + 81 \\ 105 &= 1 \cdot 81 + 24 \\ 81 &= 3 \cdot 24 + 9 \\ 24 &= 2 \cdot 9 + 6 \\ 9 &= 1 \cdot 6 + 3 \\ 6 &= 2 \cdot 3 + 0 \end{aligned}$$

So $\gcd(1665, 978) = 3$.

The algorithm

The input for the algorithm is an ordered pair of numbers and this pair is replaced by a pair having as first number the second of the previous pair. A descending sequence of numbers is obtained with 0 as last number and the greatest common divisor as the last but one:

1665	987	687	291	105	81	24	9	6	3	0

Python

We start a new python module, `arithmetics.py`. In this module we use the functions and methods in Python for the data type `integer`. We do not restrict any more to `succ` as we did in the module `integer`.

```
>>> 34 + 72                              #sum
106
>>> 34 - 72                              #difference
-38
>>> 45 * 12                              #product
540
>>> 89 // 7                              #quotient
12
>>> 89 % 7                               #remainder
5
>>> divmod(89, 7)                        #quotient,remainder
(12, 5)
>>> pow(5, 8)                            #exponentiation
390625
>>> 5**8                                 #exponentiation
390625
```

The Euclidean algorithm is easily programmed in Python. The function `gcd` as defined below returns the greatest common divisor using this algorithm. It is the first function in the module `arithmetics.py`.

arithmetics.py

```
def gcd(a, b):
    while b > 0: a, b = b, a % b
    return a
```

```
>>> gcd(1665, 987)
3
>>> gcd(12578898999900224988,34536788999000823540)
12
```

We represent rational numbers as 2-tuples of integers. Addition and multiplication are easily described in Python. Fractions are simplified using the Euclidean algorithm.

arithmetics.py

```
def simplify(a, b):
    d = gcd(a, b)
    return (a // d, b // d)
```

arithmetics.py

```
def add(x, y):
    return simplify(x[0] * y[1] + x[1] * y[0], x[1] * y[1])

def mul(x, y):
    return simplify(x[0] * y[0], x[1] * y[1])
```

```
>>> simplify(234513, 378)
(26057, 42)
>>> add((2444, 2511), (112, 15))
(105964, 12555)
>>> mul((-2334, 45), (3301, 4562))
(-1284089, 34215)
```

9.5 Properties of the Greatest Common Divisor

The following theorem is a property of the greatest common divisor that is fundamental for many applications.

9.32 Theorem. *Let a and b be integers, not both 0. Then there exist integers x, y such that $xa + yb = \gcd(a, b)$.*

Proof. The set

$$V = \{\, xa + yb \mid x, y \in \mathbb{Z} \text{ and } xa + yb > 0 \,\}$$

is a nonempty subset of $\mathbb{N}^+$: if for example $a > 0$, then $a = 1 \cdot a + 0 \cdot b \in V$. Because $V \neq \emptyset$, there is a least element c in V, say $c = x_0 a + y_0 b$ with $x_0, y_0 \in \mathbb{Z}$. We will prove that $c = \gcd(a, b)$.

From $\gcd(a, b) \mid a$ and $\gcd(a, b) \mid b$ it follows that $\gcd(a, b) \mid x_0 a + y_0 b = c$.

Division with remainder of a by c gives $a = qc + r$ with $r \in \mathbb{N}_c$. For r we have

$$r = a - qc = a - qx_0 a - qy_0 b = (1 - qx_0)a + (-qy_0)b.$$

For $r \neq 0$ it follows that $r \in V$, which however is not the case, because $r < c$ and c is the least in V. So $r = 0$, that is $c \mid a$. Similarly, $c \mid b$. So $c \leq \gcd(a, b)$. Since also $\gcd(a, b) \mid c$, it follows that $\gcd(a, b) = c$ □

Alternatively, it is not hard to see that the Euclidean algorithm produces the greatest common divisor as a combination of the given numbers. At the end of this section we will see how an extension of the Euclidean algorithm makes a fast computation of the x and y possible.

We derive some easy, but very useful, consequences of the above theorem.

9.33 Proposition. *Common divisors are divisors of the greatest common divisor.*

Proof.

Let c be a common divisor of a and b. We assume that $a, b \in \mathbb{N}$. Then by theorem 9.32 there are $x, y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$. From $c \mid a$ and $c \mid b$ follows $c \mid xa + yb$, that is $c \mid \gcd(a, b)$.

So every common divisor of a and b is a divisor of the greatest common divisor of a and b. □

9.34 Example. The common divisors of 24 and -18 are ± 1, ± 2, ± 3, ± 6, which are the divisors of 6, the greatest common divisor of 24 and -18.

9.35 Proposition. *Let a, b and c be integers. Then*

$$a \mid bc \quad \textit{and} \quad \gcd(a, b) = 1 \quad \Longrightarrow \quad a \mid c.$$

Proof.

Suppose $a \mid bc$ and $\gcd(a, b) = 1$. By theorem 9.32 there are $x, y \in \mathbb{Z}$ such that $xa + yb = 1$. Then $xac + ybc = c$. Since $a \mid xac$ and $a \mid ybc$, we also have $a \mid c$.

So: if $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$. □

9.36 Proposition. *Let a, b and c be integers with $a, b \neq 0$. Then*

$$a \mid c, \quad b \mid c \quad \textit{and} \quad \gcd(a, b) = 1 \quad \Longrightarrow \quad ab \mid c.$$

Proof. Suppose $a \mid c$, $b \mid c$ and $\gcd(a, b) = 1$. Let $d \in \mathbb{Z}$ be such that $ad = c$. From $\gcd(a, b) = 1$ and $b \mid c$ by proposition 9.35 it follows that $b \mid d$. And so $ab \mid ad$, that is $ab \mid c$. □

9.37 Proposition. *Let a and b be integers, not both 0. Let c be an integer $\neq 0$. Then*

$$\gcd(ac, bc) = \gcd(a, b) \cdot c.$$

Proof. If d is a common divisor of a and b, then dc is a common divisor of ac and bc, and so dc is a divisor of $\gcd(ac, bc)$. In particular for $d = \gcd(a, b)$ we have:

$$\gcd(a, b) \cdot c \mid \gcd(ac, bc).$$

There are integers x, y with $ax + by = \gcd(a, b)$. Then $acx + bcy = \gcd(a, b) \cdot c$, which implies:

$$\gcd(ac, bc) \mid acx + bcy = \gcd(a, b) \cdot c. \qquad \square$$

9.5.1 Reducing fractions

A rational number r is uniquely representable by a fraction having the least denominator of r as denominator. Such fractions have a positive denominator and their numerator and denominator are relatively prime. They are said to be in be *reduced* or *in reduced form.*

9.38 Proposition. *Let n be the least denominator of a rational number r. Then every denominator of r is a multiple of n.*

Proof. Since n is the least denominator of r, we have $\gcd(nr, n) = 1$, because otherwise $\frac{n}{\gcd(nr,n)}$ would be a denominator less than n. Let m be any denominator of r. Since $n \mid n \cdot mr = m \cdot nr$, it follows from proposition 9.35 that $n \mid m$. $\square$

9.5.2 Linear Diophantine equations

We will use properties of the greatest common divisor for the solution of Diophantine equations of type

$$ax + by = c,$$

where $a, b, c \in \mathbb{Z}$. The objective is to determine all solutions $x, y \in \mathbb{Z}$.

9.39 Theorem. *Let a, b and c be integers with a and b not both 0. Then the Diophantine equation*

$$ax + by = c$$

is solvable if and only if $\gcd(a, b) \mid c$.

Proof.

Suppose the equation is solvable and $(x_0, y_0) \in \mathbb{Z}^2$ is a solution, that is

$$ax_0 + by_0 = c.$$

Then, since $\gcd(a, b) \mid a$ and $\gcd(a, b) \mid b$, we also have $\gcd(a, b) \mid c$.

So: if the equation is solvable, then $\gcd(a, b) \mid c$.

Diophantus ((probably) Alexandria 200 – 284)

Diophantine equations are equations where integer solutions are asked for. They are called that way after **Diophantus**. He wrote a series of 13 books called *Arithmetica* about the solution of algebraic equations and number theory. Only six books survived and probably there are traces of the other books in Arabic texts.

Suppose $\gcd(a,b) \mid c$, say $c = k \cdot \gcd(a,b)$ with $k \in \mathbb{Z}$. From theorem 9.32 it follows that there are numbers $u, v \in \mathbb{Z}$ such that

$$au + bv = \gcd(a,b).$$

Then also

$$a(ku) + b(kv) = k \cdot \gcd(a,b)\ (= c)$$

and so (ku, kv) is a solution of $ax + by = c$.

So: if $\gcd(a,b) \mid c$, then the equation is solvable. □

9.40 Example. The Diophantine equation

$$36x + 21y = -1$$

is not solvable, because $\gcd(36,21) = 3 \nmid -1$. On the other hand the Diophantine equation

$$36x + 21y = 24$$

is solvable, because $\gcd(36,21) \mid 24$. Having a solution of $36x + 21y = 3$, a solution of $36x + 21y = 24$ is obtained by multiplying with 8. One of the solutions of $12x + 7y = 1$ is $(3,-5)$, which is also a solution of $36x + 21y = 3$; so a solution of $36x+21y = 24$ is $(24,-40)$. We will determine all $(x,y) \in \mathbb{Z}^2$ such that $36x+21y = 24$.

Suppose $x, y \in \mathbb{Z}$ satisfy $36x + 21y = 24$. Then

$$36(x-24) + 21(y+40) = 0,$$

that is

$$12(x-24) = -7(y+40).$$

Because $\gcd(12, -7) = 1$, proposition 9.35 implies $12 \mid (y+40)$, so there exists a $t \in \mathbb{Z}$ such that $y + 40 = 12t$, that is

$$y = -40 + 12t.$$

Then $12(x - 24) = -7 \cdot 12t$, or

$$x = 24 - 7t.$$

So the solutions of $36x + 21y = 24$ are of the form

$$\begin{cases} x = 24 - 7t \\ y = -40 + 12t \end{cases} \qquad (\text{where } t \in \mathbb{Z})$$

and all these are in fact solutions. Thus we have determined all solutions of the equation.

9.5.3 The least common multiple

9.41 Definition. Let a and b be integers. We say that $c \in \mathbb{Z}$ is a *common multiple* of a and b if

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

If $a \neq 0$ and $b \neq 0$, then there are positive common multiples of a and b. The least among them is called the *least common multiple* of a and b; notation: $\operatorname{lcm}(a, b)$. It is only defined for $a \neq 0$ and $b \neq 0$.

9.42 Example. The positive multiples of 24 are 24, 48, 72, 96, 120, etc. The least one which is also a multiple of -18 is 72. So 72 is the least common multiple of 24 and -18.

The least common multiple is closely related to the greatest common divisor:

9.43 Theorem. *Let $a, b \in \mathbb{N}^+$. Then*

$$\gcd(a, b) \cdot \operatorname{lcm}(a, b) = ab.$$

Proof. Put $c = \operatorname{lcm}(a, b)$ and $d = \gcd(a, b)$. Because $\frac{a}{d} \mid \frac{c}{d}$ and $\frac{b}{d} \mid \frac{c}{d}$, lemma 9.36 and proposition 9.37 imply that $\frac{ab}{d^2} \mid \frac{c}{d}$, that is $ab \mid cd$. From $\frac{ab}{d} = \frac{a}{d} \cdot b = a \cdot \frac{b}{d}$ it is clear that $\frac{ab}{d}$ is a common multiple of both a and b. So $c \leq \frac{ab}{d}$, or $cd \leq ab$. Together with $ab \mid cd$ this proves the theorem. □

9.44 Corollary. *Let a and b be nonzero integers. Then there are integers x and y such that*

$$\frac{1}{\operatorname{lcm}(a, b)} = \frac{x}{a} + \frac{y}{b}.$$

Proof. There are integers y and x such that $\gcd(a,b) = ya + xb$. Then by theorem 9.43

$$\frac{1}{\operatorname{lcm}(a,b)} = \frac{\gcd(a,b)}{ab} = \frac{ya+xb}{ab} = \frac{x}{a} + \frac{y}{b}. \qquad \square$$

The extended Euclidean algorithm

By lemma 9.29 the determination of numbers x and y such that $ax+by = \gcd(a,b)$ for given natural numbers a and $b \neq 0$ can be reduced to the determination of numbers u and v such that $bu + rv = \gcd(b,r)(= \gcd(a,b))$, where $a = qb + r$ and $r < b$. By keeping how each newly calculated remainder in the Euclidean algorithm is a combination of the original pair of numbers a, b, we finally obtain such a combination for the greatest common divisor. As an example we do this for 65 and 23:

65	23	19	4	3	1	0
		2	1	4	1	3
1	0	1	-1	5	-6	23
0	1	-2	3	-14	17	-65

We proceed from left to right. In the second row we put the quotients, and above these the new remainder: 19 is obtained by subtracting 2×23 from 65. In the two rows at the bottom every new number is obtained in the same way as a combination of the two preceding numbers. In this example the greatest common divisor is 1. Should we have started with $65a$ and $23a$, the numbers in the top row would have been multiplied by a.

Python

arithmetics.py

```
def euclid(a, b):
    p = q = y = u = 0
    x = v = 1
    while b > 0:
        p, a, q, b = (q, b) + divmod(a, b)
        x, y, u, v = u, v, x - q * u, y - q * v
    return x, y, a
```

```
>>> euclid(23414455667700999121, 19988778383883201442)
(-423838664099439595, 496476143775527043, 11)
```

9.6 Finite Continued Fractions

The Euclidean algorithm applied to 65 and 23 goes as follows:

$$65 = 2 \cdot 23 + 19 \qquad \text{or also:} \qquad \frac{65}{23} = 2 + \frac{19}{23}$$

$$\begin{aligned} 23 &= 1 \cdot 19 + 4 \\ 19 &= 4 \cdot 4 + 3 \\ 4 &= 1 \cdot 3 + 1 \\ 3 &= 3 \cdot 1 + 0 \end{aligned} \qquad \begin{aligned} \tfrac{23}{19} &= 1 + \tfrac{4}{19} \\ \tfrac{19}{4} &= 4 + \tfrac{3}{4} \\ \tfrac{4}{3} &= 1 + \tfrac{1}{3} \end{aligned}$$

By the same calculation:

$$\frac{65}{23} = 2 + \frac{19}{23} = 2 + \cfrac{1}{1 + \cfrac{4}{19}} = 2 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{3}{4}}} = 2 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{1 + \cfrac{1}{3}}}}.$$

The result is a so-called continued fraction.

9.45 Definition. We define the *continued fraction* $\langle x_1, \dots, x_n \rangle$ of a finite sequence of rational (later, in chapter 17: real) numbers $x_1, \dots, x_n$ with $x_2, \dots, x_n > 0$ inductively:

$$\begin{cases} \langle x_1 \rangle = x_1 \\ \langle x_1, x_2 \rangle = x_1 + \dfrac{1}{x_2} \\ \langle x_1, \dots, x_n \rangle = \langle x_1, \langle x_2, \dots, x_{n-2}, x_{n-1}, x_n \rangle \rangle \quad \text{for all } n \geq 3. \end{cases}$$

Thus the notation $\langle x_1, x_2, \dots, x_n \rangle$ is in fact short for $\langle x_1, \langle x_2, \langle x_3, \dots, \langle x_n \rangle \dots \rangle \rangle \rangle$. For numbers x_1, x_2, x_3, x_4 with $x_2, x_3, x_4 > 0$ we have:

$$\langle x_1, x_2, x_3 \rangle = \langle x_1, \langle x_2, x_3 \rangle \rangle = x_1 + \frac{1}{\langle x_2, x_3 \rangle} = x_1 + \cfrac{1}{x_2 + \cfrac{1}{x_3}}.$$

$$\langle x_1, x_2, x_3, x_4 \rangle = x_1 + \frac{1}{\langle x_2, x_3, x_4 \rangle} = x_1 + \cfrac{1}{x_2 + \cfrac{1}{x_3 + \cfrac{1}{x_4}}}.$$

Using continued fractions the Euclidean algorithm takes the form:

$$\tfrac{65}{23} = \langle 2, \tfrac{23}{19} \rangle = \langle 2, 1, \tfrac{19}{4} \rangle = \langle 2, 1, 4, \tfrac{4}{3} \rangle = \langle 2, 1, 4, 1, 3 \rangle.$$

Note that also $\frac{65}{23} = \langle 2, 1, 4, 1, 2, 1 \rangle$, because $3 = \langle 2, 1 \rangle$.

We write a few continued fractions as ordinary fractions, that is with only one division bar:

$$\langle x_1 \rangle = x_1 = \frac{x_1}{1}$$

$$\langle x_1, x_2\rangle = x_1 + \frac{1}{x_2} = \frac{x_2x_1+1}{x_2}$$
$$\langle x_1, x_2, x_3\rangle = x_1 + \cfrac{1}{x_2 + \cfrac{1}{x_3}} = \frac{x_3x_2x_1 + x_3 + x_1}{x_3x_2 + 1}.$$

The numerator and the denominator are in these cases polynomials in $x_1, x_2, \ldots$. Below we will define polynomials $p_n(x_1, \ldots, x_n)$ and $q_n(x_1, \ldots, x_n)$ and subsequently we will prove that indeed $\langle x_1, \ldots, x_n\rangle = \frac{p_n(x_1,\ldots,x_n)}{q_n(x_1,\ldots,x_n)}$.

9.46 Definition. We define the sequence of polynomials $p_{-1}, p_0, p_1(x_1), p_2(x_1, x_2)$, $p_3(x_1, x_2, x_3), \ldots$ by

$$\begin{cases} p_{-1} = 0 \\ p_0 = 1 \\ p_n(x_1, \ldots, x_n) = x_n p_{n-1}(x_1, \ldots, x_{n-1}) + p_{n-2}(x_1, \ldots, x_{n-2}) \quad (\text{all } n \in \mathbb{N}^+). \end{cases}$$

The Euclidean algorithm applied to 65 and 23 gives $\frac{65}{23} = \langle 2, 1, 4, 1, 3\rangle$. With the use of the polynomials p_n the fraction $\frac{65}{23}$ can be retrieved:

$$\begin{aligned} p_{-1} &= 0 \\ p_0 &= 1 \\ p_1(3) &= 3 \cdot p_0 + p_{-1} = 3 \\ p_2(3, 1) &= 1 \cdot p_1(3) + p_0 = 4 \\ p_3(3, 1, 4) &= 4 \cdot p_2(3, 1) + p_1(3) = 19 \\ p_4(3, 1, 4, 1) &= 1 \cdot p_3(3, 1, 4) + p_2(3, 1) = 23 \\ p_5(3, 1, 4, 1, 2) &= 2 \cdot p_4(3, 1, 4, 1) + p_3(3, 1, 4) = 65 \end{aligned}$$

For this method the quotients occurring in the algorithm are needed in reverse order. Thus the remainders in the algorithm are retrieved in reverse order as well. Remarkably enough it can be done the other way round. Though this might be proved directly, we will derive it from the following theorem.

9.47 Theorem. *For every $n \in \mathbb{N}^+$ we have $p_n(x_1, \ldots, x_n) = p_n(x_n, \ldots, x_1)$.*

Proof. Let $P(n)$ be the proposition

$$p_n(x_1, \ldots, x_n) = p_n(x_n, \ldots, x_1).$$

We will prove by mathematical induction that $P(n-1)$ and $P(n)$ hold for all $n \geq 2$. For $n = 2$ it is obviously true.

Suppose $P(n-1)$ and $P(n)$ hold for some $n \in \mathbb{N}$ with $n \geq 2$. Then to prove that also $P(n+1)$ holds.

$$\begin{aligned}
p_{n+1}(x_1,\dots,x_{n+1}) &= x_{n+1}p_n(x_1,\dots,x_n) + p_{n-1}(x_1,\dots,x_{n-1}) \\
&= x_{n+1}p_n(x_n,\dots,x_1) + p_{n-1}(x_{n-1},\dots,x_1) \\
&= x_{n+1}x_1p_{n-1}(x_n,\dots,x_2) + x_{n+1}p_{n-2}(x_n,\dots,x_3) \\
&\quad + x_1p_{n-2}(x_{n-1},\dots,x_2) + p_{n-3}(x_{n-1},\dots,x_3) \\
&= x_1(x_{n+1}p_{n-1}(x_n,\dots,x_2) + p_{n-2}(x_{n-1},\dots,x_2)) \\
&\quad + x_{n+1}p_{n-2}(x_n,\dots,x_3) + p_{n-3}(x_{n-3},\dots,x_3) \\
&= x_1(x_{n+1}p_{n-1}(x_2,\dots,x_n) + p_{n-2}(x_2,\dots,x_{n-1})) \\
&\quad + x_{n+1}p_{n-2}(x_3,\dots,x_n) + p_{n-3}(x_3,\dots,x_{n-1}) \\
&= x_1p_n(x_2,\dots,x_{n+1}) + p_{n-1}(x_3,\dots,x_{n+1}) \\
&= x_1p_n(x_{n+1},\dots,x_2) + p_{n-1}(x_{n+1},\dots,x_3) \\
&= p_{n+1}(x_{n+1},\dots,x_1). \qquad \square
\end{aligned}$$

9.48 Corollary. *For every $n \in \mathbb{N}^+$ we have*

$$p_n(x_1,\dots,x_n) = x_1p_{n-1}(x_2,\dots,x_n) + p_{n-2}(x_3,\dots,x_n).$$

Proof.

$$\begin{aligned}
p_n(x_1,\dots,x_n) &= p_n(x_n,\dots,x_1) = x_1p_{n-1}(x_n,\dots,x_2) + p_{n-2}(x_n,\dots,x_3) \\
&= x_1p_{n-1}(x_2,\dots,x_n) + p_{n-2}(x_3,\dots,x_n). \qquad \square
\end{aligned}$$

9.49 Definition. The sequence of polynomials $q_{-1}, q_0, q_1(x_1), q_2(x_1,x_2),\dots$ is defined by $q_{-1} = 1$ and $q_n(x_1,\dots,x_n) = p_{n-1}(x_2,\dots,x_n)$ for $n \geq 0$.

Thus the sequence of polynomials q_n is constructed the same way as the sequence of the p_n, only the initial values (for $n = -1$ and $n = 0$) are not 0 and 1, but 1 and 0.

9.50 Theorem. *For all $n \in \mathbb{N}^+$ we have*

$$\langle x_1,\dots,x_n\rangle = \frac{p_n(x_1,\dots,x_n)}{q_n(x_1,\dots,x_n)}$$

for all rational (later: real) numbers $x_1,\dots,x_n$ with $x_2,\dots,x_n > 0$.

Proof. We prove this by induction on n. For $n = 1$ it is clearly true.

Suppose it holds for some $n \in \mathbb{N}^+$. Then by Corollary 9.48:

$$\langle x_1,\dots,x_{n+1}\rangle = x_1 + \frac{1}{\langle x_2,\dots,x_{n+1}\rangle} = x_1 + \frac{q_n(x_2,\dots,x_{n+1})}{p_n(x_2,\dots,x_{n+1})}$$

$$
\begin{aligned}
&= x_1 + \frac{p_n(x_3, \ldots, x_{n+1})}{p_n(x_2, \ldots, x_{n+1})} \\
&= \frac{x_1 p_n(x_2, \ldots, x_{n+1}) + p_{n-1}(x_3, \ldots, x_{n+1})}{p_n(x_2, \ldots, x_{n+1})} \\
&= \frac{p_{n+1}(x_1, \ldots, x_{n+1})}{p_n(x_2, \ldots, x_{n+1})} = \frac{p_{n+1}(x_1, \ldots, x_{n+1})}{q_{n+1}(x_1, \ldots, x_{n+1})}.
\end{aligned}
$$

□

9.51 Example. We rewrite $\langle 2, 1, 4, 1, 3\rangle$ as an ordinary fraction using the above theorem:

i:	-1	0	1	2	3	4	5
x_i:			2	1	4	1	3
p_i:	0	1	2	3	14	17	65
q_i:	1	0	1	1	5	6	23

So: $\langle 2, 1, 4, 1, 3\rangle = \frac{p_5}{q_5} = \frac{65}{23}$. (The p_i and q_i are calculated from left to right: for example $p_5 = 3 \cdot 17 + 14 = 65$ and $q_3 = 4 \cdot 1 + 1 = 5$.)

Compare this with the extended Euclidean algorithm. Apart from the sign the same numbers occur, namely $(-1)^{n+1}p_n$ and $(-1)^n q_n$. This is easily proved by induction.

9.52 Theorem. *For all $n \geq -1$ we have*

$$p_n q_{n+1} - p_{n+1} q_n = (-1)^n,$$

for all numbers $x_1, \ldots, x_{n+1}$ with $x_2, \ldots, x_{n+1} > 0$, where p_k stands for $p_k(x_1, \ldots, x_k)$ and analogously for q_k.

Proof. The proof is by mathematical induction. For $n = -1$ the formula holds:

$$p_{-1}q_0 - p_0 q_{-1} = 0 \cdot 0 - 1 \cdot 1 = -1 = (-1)^{-1}.$$

Let $n \geq -1$ such that $p_n q_{n+1} - p_{n+1} q_n = (-1)^n$ for all $x_1, \ldots, x_{n+1}$ such that $x_2, \ldots, x_{n+1} > 0$. Suppose $x_1, \ldots, x_{n+2}$ are numbers such that $x_1, \ldots, x_{n+2} > 0$. Then

$$
\begin{aligned}
p_{n+1}q_{n+2} - p_{n+2}q_{n+1} &= p_{n+1}(x_{n+1}q_{n+1} + q_n) - (x_{n+1}p_{n+1} + p_n)q_{n+1} \\
&= -(p_n q_{n+1} - p_{n+1} q_n) \\
&= -(-1)^n = (-1)^{n+1}.
\end{aligned}
$$

□

9.53 Corollary. *Let $n \in \mathbb{N}^+$ and let $a_1, \ldots, a_n$ be integers with $a_2, \ldots, a_n \in \mathbb{N}^+$. Then*

$$\gcd(p_n(a_1, \ldots, a_n), q_n(a_1, \ldots, a_n)) = 1.$$

Proof. We write p_k for $p_k(a_1, \ldots, a_k)$ and analogously for q_k. Note that $p_k, q_k \in \mathbb{Z}$ for all $k \geq -1$. From

$$p_{n-1}q_n - q_{n-1}p_n = (-1)^{n-1}$$

it follows that 1 is the only positive common divisor of p_n and q_n. So their greatest common divisor is 1. □

So for $a_1 \in \mathbb{Z}$ and $a_2, \ldots, a_n \in \mathbb{N}^+$ we have

$$\langle a_1, \ldots, a_n \rangle = \frac{p_n}{q_n},$$

with $\gcd(p_n, q_n) = 1$. Writing for example $\frac{130}{46}$ as such a continued fraction we find $\frac{130}{46} = \langle 2, 1, 4, 1, 3 \rangle$ and so $\frac{130}{46} = \frac{p_5}{q_5}$ with $p_5 = 65$ and $q_5 = 23$.

Python

Conversion of fractions into continued fractions and vice versa is easily done by computer. The conversion from ordinary fraction to continued fraction can be supplemented to the extended Euclidean algorithm as described on page 156.

arithmetics.py

```
def confrac(a, b):
  c = ()
  while b > 0:
    d = divmod(a, b)
    a, b, c = b, d[1], c + (d[0], )
  return c

def fract(con):
    r, p = 0, 1
    s, q = 1, 0
    for i in con:
        r, p = p, r + i * p
        s, q = q, s + i * q
    return (p, q)

def euclidext(a, b):
    i = 0
    r, p = 0, 1
    s, q= 1, 0
    while b > 0:
        i = i + 1
        d = divmod(a, b)
        r, p = p, d[0] * p + r
        s, q = q, d[0] * q + s
        a, b = b, d[1]
    return (a, (-1)**i * s, -(-1)**i * r)
```

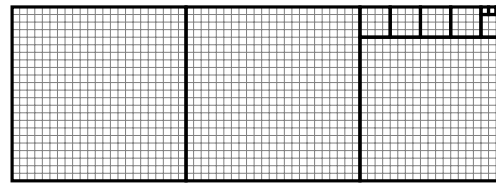
```
>>> confrac(240967887, 15570982)
(15, 2, 9, 1, 2, 7, 7, 16, 1, 2, 2, 13, 3)
>>> fract((1, 2, 3, 4, 5, 6, 7, 8, 9, 10))
(7489051, 5225670)
>>> euclidext(2364654, 74637435)
(3, -10520368, 333305)
```

9.7 Geometry and Rational Numbers

The Euclidean algorithm for the determination of the greatest common divisor of two natural numbers has a geometrical interpretation. A line segment a is called a *measure* for a line segment b if a goes into b exactly an integral number of times. Two line segments are called *commensurable* if they have a common measure. The geometrical version of the Euclidean algorithm is used for determining the greatest common measure of two commensurable line segments. That is the way the Euclidean algorithm was considered in ancient Greece. The focus was on line segments and not on numbers.

The same process can be applied to two *incommensurable* line segments, but in that case it does not terminate. On the other hand it returns better and better approximations to the ratio of the two line segments.

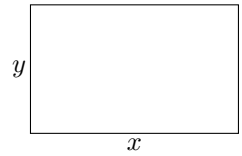
With the Euclidean algorithm integral multiples of a line segment is subtracted from another line segment. Taking the two line segments as the sides of a rectangle, the process can be seen as the subtraction of squares from the rectangle as often as possible. Finally the side of a small square in the figure is the the greatest common measure of the sides of the rectangle. Applied to a rectangle with sides of length 65 and 23 the result is:



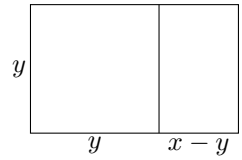
The Golden Ratio

In ancient Greece already there was much interest in a special ratio, the *golden ratio*. It is the ratio obtained by dividing a line segment into two parts, a line segment of length x and a smaller line segment of length y such that the ratio of x and y equals the ratio of $x + y$ and x.

This ratio was considered to be the ideal ratio for the sides of a rectangle, the *golden rectangle*:



A golden rectangle is characterized by the fact that after subtraction of a square as in the figure below, a similar rectangle remains:



Put $\tau = \frac{x}{y}$. Then from

$$\frac{y}{x-y} = \frac{x}{y}$$

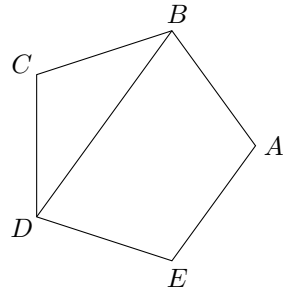
it follows that

$$\tau = \frac{x}{y} = \frac{1}{\frac{x-y}{y}} = \frac{1}{\tau - 1},$$

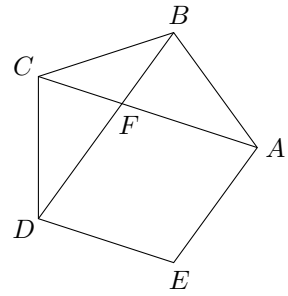
that is $\tau^2 = \tau + 1$. For the existence of a solution of this quadratic equation 5 has to be a square. In chapter 10 we will see that this is not the case in $\mathbb{Q}$.

Another argument is: suppose $\tau = \frac{a}{b}$ with $a, b \in \mathbb{N}^+$, then: $\frac{a}{b} = \frac{b}{a-b}$. Since $b < a$ we have written τ with a numerator less than a. Repeating this results in a strictly descending sequence of numerators in $\mathbb{N}^+$. Such a sequence does not exist.

The golden ratio τ is the ratio of a diagonal and a side of a regular pentagon.



It is not hard to see this. Just draw an extra diagonal.



The diagonal BD is parallel to the side AE and the diagonal AC to the side DE. So $DEAF$ is a diamond. The triangle BCD is proportional to the triangle CFB. Let x be the length of a diagonal and y the length of a side. Then

$$\frac{x}{y} = \frac{y}{x-y}.$$

Exercises

1. What goes wrong in the construction of $\mathbb{Q}$ if we use all pairs $(a, b) \in \mathbb{Z}^2$, so without the condition $b \neq 0$?
2. Prove that the set $\mathbb{Q}$ is countable. (Hint: exercises 12, 13 and 14 of chapter 5.)
3. Let r and s be rational numbers with $r < s$. Show that there are infinitely many rational numbers t with $r < t < s$.
4. Verify that $x = 4$ is a solution of $9x^3 = 109x + 140$. Are there other solutions? How many? Which?
5. Let a and b be rational numbers. Let $x = b$ be a solution of the equation $x^3 = a$. Show that it is the only solution (in $\mathbb{Q}$).
6. Show that the relation $|$ ('is a divisor of') in $\mathbb{N}$ is an ordering of $\mathbb{N}$.
7. Simplify $\frac{1207}{595}$ and $\frac{222677}{-574469}$.
8. Let a and b be odd integers such that $\gcd(a, b) = 1$. Show that
$$\gcd(a + b, a - b) = 2.$$
9. (i) Use the Euclidean algorithm to determine the greatest common divisor of
   a) 45 and 75,

b) 102 and 442,

c) 1616 and 444,

d) 87505 and 23445.

(ii) Write in each of these cases the greatest common divisor as a combination of the two integers.

10. Determine the greatest common divisor of 111111 and 11111111. What is the greatest common divisor of $\overbrace{111\ldots111}^{m}$ and $\overbrace{111\ldots111}^{n}$?

11. Determine all $x, y \in \mathbb{Z}$ satisfying

$$17x + 12y = 21.$$

12. A regular n-gon is called *constructible* if it can be constructed with a straight edge and a compass starting from its center and one of its vertices.

(i) Let m and n be natural numbers with $m, n \neq 0$ and $m \mid n$. Show: if a regular n-gon is constructible, then so is a regular m-gon.

(ii) Show: if a regular n-gon is constructible, then so is a regular $2n$-gon.

(iii) Let m and n be natural numbers ≥ 3 such that $\gcd(m, n) = 1$. Show: if regular m- and n-gons are constructible, then so is a regular mn-gon.

13. Write $\frac{460}{267}$ as a continued fraction of integers.

14. Write $\langle 2, 2, 2, 2, 2, 2, 2, 2\rangle$ as an ordinary fraction.

15. Let $x_1, \ldots x_n$ be rational numbers with $x_2, \ldots, x_n > 0$. Prove that

$$p_n(x_1, \ldots, x_n) = \langle x_1, \ldots, x_n\rangle \cdot \langle x_2, \ldots, x_n\rangle \cdot \langle x_3, \ldots, x_n\rangle \cdots \langle x_{n-1}, x_n\rangle \cdot \langle x_n\rangle.$$

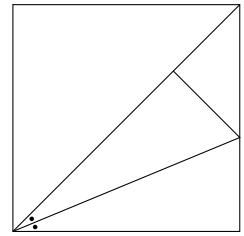
16. Let $a_1 \in \mathbb{Z}$ and $a_2, \ldots, a_n \in \mathbb{N}^+$, where $n \geq 3$. Prove that $\lfloor\langle a_1, \ldots, a_n\rangle\rfloor = a_1$.

17. Let A be the set of all finite sequences of length > 0 in $\mathbb{N}^+$. The map $F\colon A \to \mathbb{Q}$ is defined by

$$F(a_1, \ldots, a_n) = \langle a_1, \ldots, a_n\rangle.$$

What is the image of F? Is F injective?

18. Use the following figure to show that there is no rational number having 2 as its square. Give a geometrical and an algebraic proof (as we did for the golden ratio).



19. We determine the rectangles with integral width x and integral length y having equal circumference and area:

$$2x + 2y = xy.$$

(i) Show that at least one of the numbers x and y is even. Let us assume that y is even: $y = 2z$ with $z \in \mathbb{N}^+$. Simplify the equation to

$$x(z-1) = 2z.$$

(ii) Show that $z - 1 \mid 2$. Use this for the determination of all solutions.

20. Let d be the greatest common divisor of the integers a and b.

(i) Let the integers u and v satisfy $ua + vb = d$. Determine integers x and y (depending on a, b, u, v) such that $xa^2 + yb = d^2$.

(ii) Derive from the previous part that $\gcd(a^2, b) \mid d^2$.

21. (i) Determine $\gcd(2^{120} - 1, 2^{75} - 1)$.

(ii) Determine $\gcd(2^{1120} - 2^{1000}, 2^{175} - 2^{100})$.

22. (i) Determine all $(x, y) \in \mathbb{Z}^2$ which satisfy

$$\frac{x}{8} + \frac{y}{125} = \frac{1}{1000}.$$

(ii) Show that $\gcd(x, y) = 1$ for all these pairs (x, y).

23. Which of the rational numbers

$$\langle 4, 1, 2, 7, 3, 8, 2\rangle, \quad \langle 4, 1, 2, 7, 3, 8, 3\rangle \quad \text{and} \quad \langle 4, 1, 2, 7, 3, 8, 2, 3\rangle$$

is the greatest? Which is the least?

24. (i) Prove that for all $n \in \mathbb{N}$

$$\gcd(2^{n+2} - 1, 2^n - 1) \mid 3.$$

(ii) Prove that for all $n \in \mathbb{N}$

$$\gcd(2^{n+2} - 1, 2^n - 1) = 3 \iff n \text{ is even.}$$

25. The relation $\equiv$ in $\mathbb{Q}$ is defined by

$$r \equiv s \iff r - s \in \mathbb{Z} \qquad (\text{for all } r, s \in \mathbb{Q}).$$

(i) Show that $\equiv$ is an equivalence relation.

(ii) Describe the equivalence class $[\frac{1}{2}]_{\equiv}$.

(iii) Give a system of representatives of the partition $\mathbb{Q}/\equiv$.

(iv) Show that under the map $f\colon \mathbb{Q} \to \mathbb{Q}/\equiv,\ r \mapsto [2r]_{\equiv}$ elements have the same image if they represent the same equivalence class.

26. The sequence $d_0, d_1, d_2, \ldots$ is given by

$$\begin{cases} d_0 = 1 \\ d_1 = 1 \\ d_{n+2} = \dfrac{d_n d_{n+1}}{d_n + d_{n+1}} \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

How are the numbers d_n related to the Fibonacci numbers?

27. We have seen that the golden ratio τ is not a rational number. It is a proof by contradiction. Show that a contradiction is also obtained when converting the fraction to a continued fraction.

Part III

Investigations and Applications

In constructing the number system we have reached in Part II the field $\mathbb{Q}$ of rational numbers. This field was obtained in two simple steps. In the next parts we extend the number system further: real numbers in Part IV and complex numbers in Part V. In this part we investigate what we have got so far. The number field system is a creation of our mind, it is imaginary. Mathematics now becomes doing research in an imaginary world. New concepts emerge, such as the concept of *prime number*. It had no role in the construction, but when investigating the multiplicative structure it is unavoidable. It leads to the Fundamental Theorem of Arithmetic (in chapter 10).
In chapter 11 we apply our knowledge of the number system to counting problems. General formulas are obtained, e.g. a formula for the number of subsets with k elements of a set with n elements.
In chapter 12 more counting problems are studied: counting problems related to permutations of finite sets. Also the structure of permutations is investigated.
Chapter 13 is about modular arithmetic. It's a kind of arithmetic with integers in which multiples of a given number are ignored. In chapter 14 arithmetic modulo a prime number is studied and also applied. The main theorem is the so-called *Quadratic Reciprocity Law*. It is applied to *prime* tests in chapter 15. The last chapter contains an application in the information technology: a widely used cryptosystem using modular arithmetic.
For an understanding of the construction of the number system most of the mathematics in this part is not needed. On the other hand, in my opinion, it belongs to the general knowledge of a mathematician, though not all mathematicians share this opinion as far the chapters 14 and 15 are concerned.

In this book chapter 14 is only needed for the computations in chapter 15 and for the application given in chapter 20.

10 The Fundamental Theorem of Arithmetic

This chapter is about the structure of $\mathbb{N}^+$ and the closely related structure of $\mathbb{Q}^*$, the structure of the abelian group of the nonzero rational numbers under multiplication. The building blocks of the monoid $\mathbb{N}^+$ are the prime numbers: every positive integer is a product of a number of primes and is so in a unique way; this is the Fundamental Theorem of Arithmetic. The prime numbers also act as building blocks for the group of positive rational numbers under multiplication.

This knowledge will be used for understanding which of the natural (and rational) numbers are n-th powers for a given natural number n. If a number is an n-th power, then obviously that number has an n-th root. Extracting roots inside $\mathbb{Q}$ is very limited, only the obvious n-th powers do have n-th roots.

In the previous chapter we considered linear Diophantine equations. Here we will consider some nonlinear ones, especially the equation $x^2 + y^2 = z^2$. The solutions of this equation are called Pythagorean triples. Geometrically this is about right triangles having sides of integer length. A well-known solution is the triple $(3, 4, 5)$.

Arithmetic functions are functions defined on $\mathbb{N}^+$. Examples are: the number of divisors, the sum of the divisors. We will apply the Fundamental Theorem of Arithmetic in the study of these functions.

10.1 Prime Factorizations

10.1 Definition. A natural number $p > 1$ is called a *prime number* (or a *prime*) if 1 and p are the only positive divisors of p. A prime number which divides an integer a is called a *prime divisor* of a. The other integers > 1 are called *composite numbers*. A divisor d of an integer a is called a *proper* divisor if $d \neq \pm a$ and $d \neq \pm 1$. So a prime number is a natural number $\neq 1$ without proper divisors.

> Apart from 0 and 1 there are two kinds of natural numbers: prime numbers and composite numbers. A composite number a has a factorization $a = bc$ with b and c natural numbers satisfying $1 < b, c < a$.

We will show that every $n \in \mathbb{N}^+$ is a product of a number of primes, where one allows no primes or only one.

	number of prime divisors:	number of prime factors:
1	0	0
2	1	1
$2^3 \cdot 3^2$	2	5
41^7	1	7

We will use the $\prod$-notation for multiplication in the same manner as the $\sum$-notation for addition. Just as the empty sum was defined to be 0, the neutral element for addition, the empty product will be 1, the neutral element for multiplication. A *prime factorization* of an $n \in \mathbb{N}^+$ can be written as follows:

$$n = \prod_p p^{k_p},$$

where the k_p are natural numbers and the product is taken over all primes. All but a finite number of the k_p are 0. For example $1665 = 3 \cdot 555 = 3 \cdot 5 \cdot 111 = 3^2 \cdot 5 \cdot 37$, and so

$$1665 = \prod_p p^{k_p}$$

with $k_3 = 2$, $k_5 = 1$, $k_{37} = 1$ and $k_p = 0$ for all $p \neq 3, 5, 37$. The product is meant to be a finite product: it is the product of all factors $\neq 1$ of which there are only finitely many.

10.2 Proposition. *Every natural number > 0 has a prime factorization.*

Proof. Let n be a natural number > 0. If $n = 1$, we are finished. For $n > 1$ let p_1 be the least divisor of n greater than 1. Note that there are divisors > 1, for example n itself is such a divisor. The least divisor cannot have proper divisors, because such divisors would be divisors of n as well. So p_1 is a prime divisor. Put $n = p_1 n_1$. Repeat this with n_1 instead of n: if $n_1 = 1$, then we are finished. Otherwise $n_1 = p_2 n_2$ with p_2 the least divisor > 1 of n_2. Thus we obtain

$$\begin{aligned}
n &= p_1 n_1 && \text{with } p_1 \text{ a prime number and } n_1 > 1,\\
n_1 &= p_2 n_2 && \text{with } p_2 \text{ a prime number and } n_2 > 1,\\
n_2 &= p_3 n_3 && \text{with } p_3 \text{ a prime number and } n_3 > 1,\\
&\vdots\\
n_{r-1} &= p_r n_r && \text{with } p_r \text{ a prime number and } n_r = 1,\\
n_r &= 1.
\end{aligned}$$

The natural numbers $n, n_1, n_2, \ldots$ satisfy $n > n_1 > n_2 > \cdots$. The process stops at an n_r with $n_r = 1$. Then $n = p_1 n_1 = p_1 p_2 n_2 = \cdots = p_1 p_2 \cdots p_r$. □

The set $\mathbb{N}^+$ together with the subset

$$\{\,(pn, n) \mid n \in \mathbb{N} \text{ and } p \text{ a prime}\,\}$$

of ${\mathbb{N}^+}^2$ is a directed graph. A walk through the graph along edges (in the indicated direction) starting in a vertex n ends in the vertex 1. Along the way the number has been divided by primes $p_1, \ldots, p_r$ and so $n = p_1 p_2 \cdots p_r$. In the above proof we divided by least prime divisors.

The prime numbers can be seen as the building blocks of the numbers in $\mathbb{N}^+$. In the next section we will prove that every such number is composed of primes in a unique way. That is what the Fundamental Theorem of Arithmetic is about. There are infinitely many natural numbers, but that does not imply that there are infinitely many building blocks. A reason why this number is infinite is already in Euclid's Elements.

10.3 Theorem (Euclid). *Let P be a finite set of prime numbers. Then there exists a prime number q with $q \notin P$.*

Proof. Let P be nonempty (otherwise take $q = 2$). Let n be the successor of the product of the primes in P:

$$n = 1 + \prod_{p \in P} p.$$

For every $p \in P$ the remainder of n after dividing by p equals 1. So $p \nmid n$ for all $p \in P$. So a prime divisor q of n is not an element of P. According to proposition 10.2 such a q exists. □

So there is no greatest prime, because otherwise the number of primes would be finite, whereas the theorem says that in that case one is missing. Having a list of the first N prime numbers, the set of the primes not in the list is nonempty and contains a least element, the $N+1$-st prime number.

10.2 The Fundamental Theorem

The Fundamental Theorem of Arithmetic states that prime factorizations are unique. In the proof we will use the following two propositions, which in fact are useful in many other occasions.

10.4 Proposition. *Let $a \in \mathbb{Z}$ with $a \neq 0$ and let p be a prime number. Then there are unique $k \in \mathbb{N}$ and $b \in \mathbb{N}^+$ such that $a = p^k b$ and $p \nmid b$.*

Proof. First we prove the existence of k and b. Consider the set S of all $m \in \mathbb{N}$ such that $p^m \mid a$. Since $p^m \leq a$ this set is finite and because $0 \in S$, it is nonempty. So it contains a greatest number k, say $a = p^k b$, where $b \in \mathbb{N}^+$. Then $p \nmid b$, because otherwise $p^{k+1} \mid a$. Each $l \in S$ different from k satisfies $l < k$ and $a = p^l p^{k-l} b$. From this the uniqueness of k (and b) follows. □

10.5 Definition. For a, p and k as in proposition 10.4 the number k is called the *p-adic value* of a. Notation: $k = v_p(a)$.

In other words $v_p(a)$ is the maximal number of factors p in a. Thus we have for each prime p a map

$$v_p \colon \mathbb{Z} \setminus \{0\} \to \mathbb{N},$$

the *p-adic valuation* on $\mathbb{Z}$. This map clearly is surjective: $v_p(p^n) = n$.

Let us denote by $\mathbb{Z}^{(p)}$ the set of integers which are not a multiple of p. Then we have a map

$$\mathbb{N} \times \mathbb{Z}^{(p)} \to \mathbb{Z},\ (k, b) \mapsto p^k b.$$

From proposition 10.4 it follows that this map is bijective, the inverse map being

$$\mathbb{Z} \to \mathbb{N} \times \mathbb{Z}^{(p)},\ a \mapsto \Big(v_p(a), \frac{a}{p^{v_p(a)}}\Big).$$

10.6 Proposition. *Let a and b be integers and p a prime. Then*

$$p \mid ab \implies p \mid a \ \text{ or } \ p \mid b.$$

Proof. Suppose $p \mid ab$. Then we must prove that $p \mid a$ or $p \mid b$. We may assume that $p \nmid a$, because otherwise we are finished. The only positive divisors of p are 1 and p, so 1 is the only positive common divisor of p and a, that is $\gcd(p, a) = 1$. From $p \mid ab$ it follows by proposition 9.35 that $p \mid b$. □

10.7 Fundamental Theorem of Arithmetic. *Prime factorizations are unique.*

Proof. Let

$$a = \prod_p p^{k_p} \quad (\text{with } k_p \in \mathbb{N}).$$

be a prime factorization of $a \in \mathbb{N}^+$. We will prove that for every prime p the exponent k_p equals $v_p(k)$.

Let q be a prime. Then $a = q^{k_q} b$ with $b = \prod_{p \neq q} p^{k_p}$. From proposition 10.6 it follows that $q \nmid b$. So by proposition 10.4 $k_q = v_q(a)$.

So for each prime p we have $k_p = v_p(a)$. □

As a consequence we can write the prime factorization of $a \in \mathbb{N}^+$ in general as follows:

$$a = \prod_p p^{v_p(a)}.$$

We are used to the uniqueness of prime factorizations to such an extent that we experience it as being obvious. The following example may show that uniqueness of factorizations is not that obvious. Let S be the set of all natural numbers having 1 as remainder after division by 4. Clearly S is closed under multiplication: if $m, n \in S$, then $mn \in S$. In S one can factorize numbers as a product of numbers without a proper factorization in S. For example $441 = 9 \cdot 49$. The numbers 9 and 49 have no proper factorization in S. However: $441 = 21 \cdot 21$ and 21 has no proper factorization in S either.

10.3 Direct Consequences of the Fundamental Theorem

Let p_n denote the n-th prime: $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \dots$. By the Fundamental Theorem of Arithmetic the map

$$\mathbb{N}^+ \to \mathcal{R}_0(\mathbb{N}), \quad a \mapsto (v_{p_1}(a), v_{p_2}(a), v_{p_3}(a), \dots)$$

is a bijection. Here $\mathcal{R}_0(\mathbb{N})$ is the set of all sequences in $\mathbb{N}$ which have a tail of zeros. Under this bijection multiplication in $\mathbb{N}^+$ corresponds to component-wise addition of sequences in $\mathbb{N}$:

10.8 Lemma. *Let a and b be nonzero natural numbers and let p be a prime number. Then:*

$$v_p(ab) = v_p(a) + v_p(b).$$

Proof. We have

$$ab = \prod_p p^{v_p(a)} \cdot \prod_p p^{v_p(b)} = \prod_p p^{v_p(a)+v_p(b)}.$$

By the Fundamental Theorem $v_p(ab) = v_p(a) + v_p(b)$ for all prime numbers p. □

This translation of the multiplication in $\mathbb{N}^+$ into the addition in $\mathcal{R}_0(\mathbb{N})$ makes the multiplicative structure of $\mathbb{N}^+$ more transparent.

10.9 Corollary. *For $a, b \in \mathbb{N}^+$ we have:*

$$a \mid b \iff v_p(a) \le v_p(b) \text{ for all primes } p.$$

Proof.

$\Rightarrow$: Suppose $a \mid b$, say $ac = b$. Then $v_p(b) = v_p(ac) = v_p(a) + v_p(c) \ge v_p(a)$ for all primes p.

$\Leftarrow$: Suppose $v_p(a) \le v_p(b)$ for all primes p. Then

$$b = a \cdot \prod_p p^{v_p(b)-v_p(a)}$$

and so $a \mid b$. □

10.10 Example. The positive divisors of 1665 are the numbers $3^i \cdot 5^j \cdot 37^k$ with $0 \le i \le 2$, $0 \le j \le 1$ and $0 \le k \le 1$; so there are $12\,(= 3 \cdot 2 \cdot 2)$ divisors. Figure 10.1 is the Hasse diagram of this ordered set (with the ordering $|$).

From the Fundamental Theorem of Arithmetic it follows that the prime factorizations of $\gcd(a, b)$ and $\operatorname{lcm}(a, b)$ are easily obtained from those of a and b:

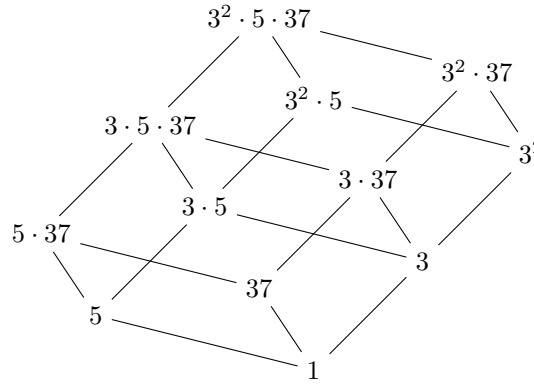


Figure 10.1: The positive divisors of 1665

10.11 Proposition. *Let $a, b \in \mathbb{N}^+$. Then for all prime numbers p:*

$$v_p(\gcd(a,b)) = \min(v_p(a), v_p(b)) \quad \textit{and} \quad v_p(\operatorname{lcm}(a,b)) = \max(v_p(a), v_p(b)).$$

Proof. A number $d \in \mathbb{N}^+$ is a common divisor of a and b if and only if $v_p(d) \leq v_p(a)$ and $v_p(d) \leq v_p(b)$, that is $v_p(d) \leq \min(v_p(a), v_p(b))$, for all prime numbers p. For the greatest common divisor we then have $v_p(\gcd(a,b)) = \min(v_p(a), v_p(b))$ for all primes p. For the least common multiple the proof is similar. □

Thus we have a way to determine the greatest common divisor of two numbers. However, since the prime factorizations of these numbers have to be determined first, this is for large numbers a formidable task. We return to this in chapter 15. The determination of the greatest common divisor is done in a very efficient way by the Euclidean algorithm, see section 9.4. Proposition 10.11 is theoretically important and for concrete calculations it is sometimes useful.

Note that the proposition gives a new and simple proof of the identity

$$\gcd(a,b) \cdot \operatorname{lcm}(a,b) = ab.$$

The Fundamental Theorem of Arithmetic describes the multiplicative structure of $\mathbb{N}^+$. It is easily extended to $\mathbb{Q}^*$.

10.12 Definition. Let p be a prime number. The *p-adic value* $v_p(r)$ of an $r \in \mathbb{Q}^*$ is defined as follows: put $r = \frac{a}{b}$ with a, b integers, then

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

Note that this does not depend on the choice of a and b: if $\frac{a}{b} = \frac{c}{d}$, then $ad = bc$ and so $v_p(a) + v_p(d) = v_p(b) + v_p(c)$, that is $v_p(a) - v_p(b) = v_p(c) - v_p(d)$.

We will denote the set of positive rational numbers by $\mathbb{Q}^{>0}$. The map

$$\mathcal{R}_0(\mathbb{Z}) \to \mathbb{Q}^{>0}, \quad (k_1, k_2, k_3, \dots) \mapsto \prod_{i \in \mathbb{N}^+} p_i^{k_i}$$

is bijective and its inverse is:

$$\mathbb{Q}^{>0} \to \mathcal{R}_0(\mathbb{Z}), \quad r \mapsto (v_{p_1}(r), v_{p_2}(r), v_{p_3}(r), \dots).$$

Multiplication (in $\mathbb{Q}^{>0}$) corresponds to component-wise addition in $\mathcal{R}_0(\mathbb{Z})$, that is $v_p(rs) = v_p(r) + v_p(s)$ for all $r, s \in \mathbb{Q}^*$ and all primes p.

Powers of Rational Numbers

The Fundamental Theorem of Arithmetic enables us to describe the rational numbers which are n-th powers (of a rational number) in terms of their prime factorizations.

Let r be a positive rational number. The prime factorization of r^n, where $n \in \mathbb{N}$, is easily determined by the prime factorization of r:

$$v_p(r^n) = n \cdot v_p(r) \text{ for all prime numbers } p.$$

So the p-adic value of an n-th power is a multiple of n for every prime p. Conversely, if the p-adic value of an $s \in \mathbb{Q}^{>0}$ is a multiple of n for every prime p, then s is an n-th power of a rational number: put $v_p(s) = k_p n$, then s is the n-th power of $r = \prod_p p^{k_p}$:

$$r^n = \prod_p p^{k_p n} = \prod_p p^{v_p(s)} = s.$$

So we have shown:

10.13 Proposition. *Let $s \in \mathbb{Q}^{>0}$ and $n \in \mathbb{N}$. Then s is an n-th power of a rational number if and only if $n \mid v_p(s)$ for every prime p. An $a \in \mathbb{N}^+$ is an n-th power of a natural number if and only if $n \mid v_p(a)$ for every prime p.* □

10.14 Examples. The number 2 is not a square of a rational number, because $v_2(2) = 1$, which is odd. Similarly, the number 5 is not a square: $v_5(5) = 1$. The rational number $\frac{27}{32}$ is not a cube of a rational number, because $v_2(\frac{27}{32}) = -5$, which is not a multiple of 3.

> By $\sqrt{2}$ we mean the positive number of which the square equals 2. Since 2 is not a square, there is no $\sqrt{2}$ in $\mathbb{Q}$. Later we will extend the rational numbers to the real numbers and then there is such a number, it just is not rational, or as one says it is irrational.

10.4 Pythagorean Triples and Fermat's Last Theorem

We will solve the Diophantine equation $x^2 + y^2 = z^2$. For the solution the Fundamental Theorem of Arithmetic will be crucial. Fermat's Last Theorem is about the solvability of the Diophantine equation $x^n + y^n = z^n$ for $n \geq 3$.

10.4.1 Pythagorean triples

10.15 Definition. A triple $(x, y, z) \in \mathbb{N}^{+3}$ is called a *Pythagorean triple* if it satisfies

$$x^2 + y^2 = z^2.$$

Examples are $(3, 4, 5)$ and $(5, 12, 13)$. Geometrically, by Pythagoras, these triples correspond to right triangles having sides of integer length.

Let (x, y, z) be a Pythagorean triple. If $d \in \mathbb{N}^+$ is a common divisor of x and y, then d is also a divisor of z, say $x = dx_0$, $y = dy_0$ and $z = dz_0$, and then (x_0, y_0, z_0) is a Pythagorean triple as well. So we can restrict our attention to so-called *primitive* Pythagorean triples, being Pythagorean triples (x, y, z) such that $\gcd(x, y) = 1$. Note that if x and y are relatively prime, the same holds for x and z, as well as for y and z.

Now let (x, y, z) be a primitive Pythagorean triple. Then x and y are not both odd, since otherwise z^2 would have 2 as the remainder after division by 4, which for squares is not the case. We will assume that x is odd and y is even. Write the equation as $z^2 - x^2 = y^2$ and factorize the left hand side:

$$(z + x)(z - x) = y^2.$$

The natural numbers $z+x$, $z-x$ and y are all even, so $\frac{z+x}{2}$, $\frac{z-x}{2}$ and $\frac{y}{2}$ are natural numbers, and we have:

$$\frac{z + x}{2} \cdot \frac{z - x}{2} = \left(\frac{y}{2}\right)^2. \tag{10.1}$$

If d is a common divisor of $\frac{z+x}{2}$ and $\frac{z-x}{2}$, then so it is of

$$x = \frac{z + x}{2} - \frac{z - x}{2} \qquad \text{and} \qquad z = \frac{z + x}{2} + \frac{z - x}{2}.$$

Since (x, y, z) is primitive, we have $\gcd(x, z) = 1$. So $\gcd(\frac{z+x}{2}, \frac{z-x}{2}) = 1$. We now use the following lemma.

10.16 Lemma. *Let $n \in \mathbb{N}^+$ and let a and b be elements of $\mathbb{N}^+$ such that ab is an nth power and $\gcd(a, b) = 1$. Then both a and b are n-th powers.*

u	v	$x = u^2 - v^2$	$y = 2uv$	$z = u^2 + v^2$
2	1	3	4	5
3	2	5	12	13
4	1	15	8	17
4	3	7	24	25
5	2	21	20	29
5	4	9	40	41
6	1	35	12	37
6	5	11	60	61
7	2	45	28	53
7	4	33	56	65
7	6	13	84	85
etc.				

Figure 10.2: Table of Pythagorean triples

Proof.

Let p be a prime divisor of a. Then $p \nmid b$ and so $v_p(a) = v_p(ab)$, which is a multiple of n.

From proposition 10.13 it follows that a is an nth power. By symmetry b is an nth power as well. □

By this lemma the identity (10.1) implies that both $\frac{z+x}{2}$ and $\frac{z-x}{2}$ are squares. Now put

$$\frac{z+x}{2} = u^2 \qquad \text{and} \qquad \frac{z-x}{2} = v^2.$$

Then $x = u^2 - v^2$, $z = u^2 + v^2$ and $y = 2uv$. So we proved:

10.17 Theorem. *Let (x, y, z) be a primitive Pythagorean triple with y even. Then there are $u, v \in \mathbb{N}^+$ such that*

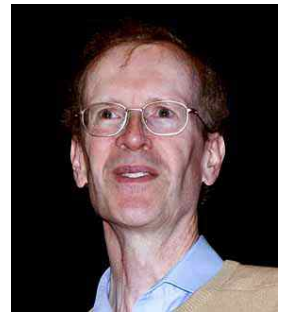
$$x = u^2 - v^2, \quad y = 2uv \quad \textit{and} \quad z = u^2 + v^2. \qquad \square$$

For which u and v is such a triple primitive? If d is a common divisor of u and v, then also of $u^2 - v^2$ and $2uv$, so $\gcd(u, v) = 1$ is needed. Furthermore, u and v cannot be both odd, because then $u^2 - v^2$ and $2uv$ would be both even, making 2 a common divisor of these numbers. These requirements suffice:

10.18 Proposition. *Let u and v be natural numbers with $u > v \geq 1$, $\gcd(u, v) = 1$ and u and v not both odd. Then $(u^2 - v^2, 2uv, u^2 + v^2)$ is a primitive Pythagorean triple.*

Andrew Wiles (Cambridge 1953)

Wiles proved Fermat's Last Theorem in 1995. He had spent seven years completing his proof. Nobody knew he was working on the conjecture, until in 1994 he presented a proof in Cambridge, but soon after a mistake was found. Luckily it could be corrected, though the correction was far from trivial.



Proof. Let d be a common divisor of u^2-v^2 and $2uv$. Then $d^2 \mid (u^2-v^2)^2 = u^4-2u^2v^2+v^4$ and $d^2 \mid 4u^2v^2$. Hence $d^2 \mid u^4+2u^2v^2+v^4 = (u^2+v^2)^2$ and so $d \mid u^2+v^2$. We have $d \mid 2u^2 (= (u^2+v^2)+(u^2-v^2))$ and $d \mid 2v^2 (= (u^2+v^2)-(u^2-v^2))$. From $\gcd(u,v) = 1$ it follows that $d \mid 2$, and since u^2+v^2 is odd, we have $d = 1$. So $\gcd(u^2-v^2, 2uv) = 1$. □

With these results it is now possible to make a potentially infinite table of all primitive Pythagorean triples, see Figure 10.2. For a different proof, see exercise 23.

10.4.2 Fermat's Last Theorem

Figure 10.3: Czech Republic, 2000

As we have seen, in infinitely many cases the sum of two squares is again a square. According to a note of Fermat in the margin of a copy of an edition of Arithmetica by Diophantus for each $n \geq 3$ the sum of two n-th powers never is an n-th power. For this he had a truly marvelous proof, but the margin was too small to contain it. This is known as Fermat's Last Theorem. Attempts to find a proof have given rise to beautiful mathematical theories. It was not proved until 1995 when the English mathematician Andrew Wiles gave a proof using techniques unavailable to Fermat. It is a consequence of a conjecture on so-called elliptic curves Wiles had proved. Fermat gave a proof for the case $n = 4$. For $n = 3$ a proof was given by Euler. In fact in his proof there was a major gap, but the necessary lemma to fill this gap can be found in Euler's work elsewhere.

Pierre de Fermat (Beaumont-de-Lomagne 1601 – Castres 1665)

Fermat was a lawyer and became well-known as an amateur mathematician, mainly because of his work in number theory. In those times it was customary to challenge each other with problems. Many results of Fermat became known because of this challenging others. The proofs were found later. 'Fermat's Last Theorem' as found in the margin of a copy of Diophantus' book was finally proved in 1995 by Wiles. In proofs like the here given proof of Theorem 10.19 Fermat used a technique known as 'infinite descent', a consequence of the well-ordering of $\mathbb{N}$.

If $d \mid n$, say $n = dm$ and (x_0, y_0, z_0) is a solution of $x^n + y^n = z^n$, then

$$(x_0^m)^d + (y_0^m)^d = (z_0^m)^d$$

and so (x_0^m, y_0^m, z_0^m) is a solution of $x^d + y^d = z^d$. Thus the problem is reduced to $n = 4$ or n an odd prime, because if n is not a multiple of an odd prime, then it is a power of 2, which is a multiple of 4 since $n > 2$.

Here we give a proof of Fermat's Last Theorem for $n = 4$. It is a consequence of the more general theorem which states that a square cannot be the sum of two fourth powers.

10.19 Theorem (Fermat). *There are no $x, y, z \in \mathbb{N}^+$ such that $x^4 + y^4 = z^2$.*

Proof.

Suppose the equation does have a solution. Let z^2 be the least square which is the sum of two fourth powers, say $z^2 = x^4 + y^4$. Then $\gcd(x, y) = 1$, since otherwise there would be a less square satisfying this property. So (x^2, y^2, z) is a primitive Pythagorean triple and therefore there are $u, v \in \mathbb{N}^+$ such that

$$x^2 = u^2 - v^2, \quad y^2 = 2uv, \quad z = u^2 + v^2$$
$$\gcd(u, v) = 1, \quad u \text{ and } v \text{ not both odd},$$

(if necessary swap x and y). Because x is odd, the remainder of x^2 after division by 4 equals 1. From $x^2 = u^2 - v^2$ it follows that u is odd and v is even. We have

$$\left(\frac{y}{2}\right)^2 = u \cdot \frac{v}{2} \qquad \text{and} \qquad \gcd\left(u, \frac{v}{2}\right) = 1,$$

so u and $\frac{v}{2}$ are both squares, say $u = a^2$ and $\frac{v}{2} = b^2$. Because (x, v, u) is a primitive Pythagorean triple, there are $c, d \in \mathbb{N}^+$ such that $\gcd(c, d) = 1$ and

$$x = c^2 - d^2, \quad v = 2cd, \quad u = c^2 + d^2.$$

Note that x is odd and v is even. From $cd = b^2$ and $\gcd(c, d) = 1$ it follows that c and d are squares. Also u is a square ($u = a^2$) and from $u = c^2 + d^2$ it now follows that u is a sum of two fourth powers, and we have

$$a^2 = u \leq u^2 < u^2 + v^2 = z \leq z^2,$$

so $a^2 < z^2$. Contradiction, since z^2 was the least square which is the sum of two fourth powers.

So no square is the sum of two fourth powers. □

The proof above shows that for every square in $\mathbb{N}^+$ which is the sum of two fourth powers there exists a smaller square with this property. So the set of all such squares does not have a least element. This contradicts Theorem 7.48: $(\mathbb{N}, \leq)$ is well-ordered. Formulated this way the proof uses the method of *infinite descent*.

10.5 Arithmetic Functions

10.20 Definition. A function $f\colon \mathbb{N}^+ \to \mathbb{Q}$ is called an *arithmetic* function.

So an arithmetic function f is just a sequence of rational numbers: $f(1)$, $f(2)$, ... Later, having more numbers at our disposal, we can extend this notion to functions taking values in $\mathbb{R}$ or $\mathbb{C}$.

10.21 Examples.

a) $\tau(n)$ = number of divisors of n. From Corollary 10.9 it follows that this number is determined by the prime factorization of n:

$$\tau(n) = \prod_p (v_p(n) + 1).$$

For example $\tau(1665) = (2+1)(1+1)(1+1) = 12$, see example 10.10.

b) $\sigma(n)$ = sum of divisors of n. For example $\sigma(1665) = 1+3+9+5+15+45+37+111+333+185+555+1665 = 2964$, see Figure 10.1.

c) $1(n) = 1$ (for all n), a constant function.

d) A function which in this context is of special importance:

$$\mathbf{1}(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \neq 1. \end{cases}$$

e) $\mathrm{id}(n) = n$ (for all n), the identity function.

Note the difference between the arithmetic functions 1, **1** and id.

Starting with an arithmetic function f a new arithmetic function F can be produced as follows:

$$F(n) = \sum_{d|n} f(d).$$

Here it is understood that the sum is over all positive divisors d of n. Examples 10.21 a) and b) are determined this way by c) and e) respectively:

$$\tau(n) = \sum_{d|n} 1 \quad \text{and} \quad \sigma(n) = \sum_{d|n} d.$$

This construction of F from f is a special case of the following.

10.22 Definition. The *Dirichlet product* $f * g$ of arithmetic functions f and g is the arithmetic function given by

$$(f * g)(n) = \sum_{d|n} f(d)g(\tfrac{n}{d}).$$

10.23 Examples.

a) $(1 * 1)(n) = \sum_{d|n} 1 \cdot 1 = \tau(n)$, so $\tau = 1 * 1$.
b) $(\mathrm{id} * 1)(n) = \sum_{d|n} d \cdot 1 = \sigma(n)$, so $\sigma = \mathrm{id} * 1$.

The Dirichlet product satisfies some simple rules:

10.24 Proposition. *The operation $*$ is associative, commutative and $\mathbf{1}$ is a neutral element.*

Proof.

Associativity:

$$\begin{aligned}((f * g) * h)(n) &= \sum_{d|n} (f * g)(d)h(\tfrac{n}{d}) = \sum_{d|n} \sum_{e|d} f(e)g(\tfrac{d}{e})h(\tfrac{n}{d}) \\ &= \sum_{\substack{(d_1,d_2,d_3) \\ d_1 d_2 d_3 = n}} f(d_1)g(d_2)h(d_3),\end{aligned}$$

where $d_1 = e$, $d_2 = \frac{d}{e}$ and $d_3 = \frac{n}{d}$. Evaluation of $(f * (g * h))(n)$ yields the same.

Commutativity: Follows directly from the definition.

Neutral element:

$$(f * \mathbf{1})(n) = \sum_{d|n} f(d)\mathbf{1}(\tfrac{n}{d}) = f(n). \qquad \square$$

The arithmetic functions together with the Dirichlet product form an abelian monoid with **1** as neutral element (unity element).

10.25 Definition. An arithmetic function f is called *multiplicative* if $f(mn) = f(m)f(n)$ for all $m, n \in \mathbb{N}^+$ with $\gcd(m,n) = 1$. It is called *strictly* multiplicative if $f(mn) = f(m)f(n)$ for all $m, n \in \mathbb{N}^+$.

10.26 Proposition. *Let f and g be multiplicative arithmetic functions. Then the function $f * g$ is also multiplicative.*

Proof. For $m, n \in \mathbb{N}^+$ with $\gcd(m,n) = 1$ we have:

$$\begin{aligned}(f*g)(mn) &= \sum_{d|mn} f(d)g(\tfrac{mn}{d}) = \sum_{\substack{d_1|m\\ d_2|n}} f(d_1d_2)g(\tfrac{m}{d_1}\cdot\tfrac{n}{d_2})\\ &= \sum_{\substack{d_1|m\\ d_2|n}} f(d_1)f(d_2)g(\tfrac{m}{d_1})g(\tfrac{n}{d_2})\\ &= \Big(\sum_{d_1|m} f(d_1)g(\tfrac{m}{d_1})\Big)\Big(\sum_{d_2|n} f(d_2)g(\tfrac{n}{d_2})\Big)\\ &= (f*g)(m)\cdot(f*g)(n).\end{aligned}$$

□

This proposition can be used to show that an arithmetic function is multiplicative. For $n \mapsto \tau(n)$ it follows from Corollary 10.9 that it is multiplicative, see also examples 10.21. Since the arithmetic function 1 clearly is multiplicative and $\tau = 1*1$, this also follows from the above proposition. Again the formula for $\tau(n)$ given in examples 10.21 follows.

10.27 Corollary. *The arithmetic function σ is multiplicative. We have:*

$$\sigma(n) = \prod_p \frac{p^{v_p(n)+1}-1}{p-1}.$$

Proof. Since the arithmetic functions τ and 1 are multiplicative, $\sigma = \tau * 1$ is multiplicative. For the power p^k of a prime number p where $k \in \mathbb{N}^+$ we have:

$$\sigma(p^k) = \sum_{d|p^k} d = \sum_{i=0}^{k} p^i = \frac{p^{k+1}-1}{p-1}.$$

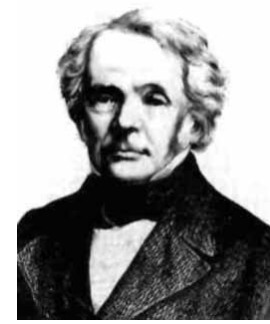
Because σ is multiplicative the formula for $\sigma(n)$ follows directly. □

So all functions in examples 10.21 are multiplicative, the functions 1, **1** and id are strictly multiplicative.

We will see that the constant arithmetic function 1 is invertible with respect to the Dirichlet product, that is it is invertible in the monoid of arithmetic functions.

August Möbius (Schulpforta 1790 – Leipzig 1868)

Möbius learned mathematics and astronomy from Gauß and Pfaff, the teacher of Gauß. He became well-known because of the Möbius strip, being an example of a one-sided surface.



10.28 Definitions. An $n \in \mathbb{N}^+$ is called *squarefree* if $v_p(n) \leq 1$ for all prime numbers p. The *Möbius function* is the arithmetic function μ defined by:

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ is not squarefree,} \\ 1 & \text{if } n \text{ is squarefree with an even number of prime divisors,} \\ -1 & \text{if } n \text{ is squarefree with an odd number of prime divisors.} \end{cases}$$

The Möbius function is the inverse of the constant function 1 in the monoid of arithmetic functions:

10.29 Lemma. $\mu * 1 = \mathbf{1}$.

Proof. The arithmetic functions μ and 1 are multiplicative. So according to proposition 10.26 the function $\mu * 1$ is multiplicative. Also $\mathbf{1}$ is multiplicative. We compute the value of the function $\mu * 1$ in prime powers. We have $(\mu * 1)(1) = \mu(1)\cdot 1 = 1$ and for p a prime number and $k \in \mathbb{N}^+$ we have $(\mu * 1)(p^k) = \sum_{d|p^k} \mu(d) = \sum_{i=0}^{k} \mu(p^i) = \mu(1) + \mu(p) = 1 - 1 = 0$. So $(\mu * 1)(n) = \mathbf{1}(n)$ for all n. □

10.30 Möbius Inversion Theorem. *Let f be an arithmetic function and let $F = f * 1$. Then $f = F * \mu$.*

Proof. $F * \mu = (f * 1) * \mu = f * (1 * \mu) = f * \mathbf{1} = f$. □

A direct consequence of this theorem and proposition 10.26 is:

10.31 Corollary. *Let f be an arithmetic function and let $F = f * 1$. Then:*

$$f \text{ is multiplicative} \iff F \text{ is multiplicative.}$$ □

Leonhard Euler (Basel 1707 – St. Petersburg 1783)

The importance of Euler for the development of mathematics has been enormous. He was by far the most productive mathematician of his time. Numerous mathematical notions bear his name. If everything he invented was named after him, his name would have been used even more. He was a master in making computations by heart. From this ability he profited when he became blind at an older age, it did not prevent him from producing new results.

10.5.1 Perfect numbers

In ancient Greece there was interest in perfect numbers for esthetic reasons. Perfect numbers are equal to the sum of their positive proper divisors. Examples are $6 = 1 + 2 + 3$ and $28 = 1 + 2 + 4 + 7 + 14$.

10.32 Definition. An $n \in \mathbb{N}^+$ is called *perfect* if $\sigma(n) = 2n$.

Euclid already gave the following type of perfect number:

10.33 Proposition. *Let $p \in \mathbb{N}$ be such that $2^p - 1$ is a prime number. Then $2^{p-1}(2^p - 1)$ is a perfect number.*

Proof. Since $\gcd(2^{p-1}, 2^p - 1) = 1$, Corollary 10.27 implies:

$$\sigma(2^{p-1}(2^p - 1)) = \frac{2^p - 1}{2 - 1} \cdot (1 + 2^p - 1) = 2^p(2^p - 1). \qquad \square$$

The numbers 6 and 28 are of this type: $6 = 2 \cdot (2^2 - 1)$ and $28 = 2^2 \cdot (2^3 - 1)$. Euler has shown that in fact all even perfect numbers are of this type:

10.34 Proposition (Euler). *Let n be an even perfect number. Then there is a $p \in \mathbb{N}^+$ such that $2^p - 1$ is a prime number and $n = 2^{p-1}(2^p - 1)$.*

Proof. Put $n = 2^{k-1}m$ with $k \geq 2$ and m odd. Then $\sigma(n) = \sigma(2^{k-1})\sigma(m) = (2^k - 1)\sigma(m)$ and, since n is perfect, also $\sigma(n) = 2n = 2^k m$. So $2^k m = (2^k - 1)\sigma(m)$. Because $\gcd(2^k - 1, 2^k) = 1$, it follows that $2^k - 1 \mid m$, say $m = (2^k - 1)q$. Then $2^k q = \sigma(m) \geq m + q = 2^k q$ and so $\sigma(m) = m + q$. So the numbers m and q are the only divisors of m. Hence $q = 1$ and m $(= 2^k - 1)$ is a prime number. $\square$

Marin Mersenne (Oize 1588 – Paris 1648)

Mersenne primes are named after the French monk Mersenne. He became known because of his work in number theory and in fact even more because of his correspondences with important mathematicians.

So the even perfect numbers are in correspondence with primes of type $2^p - 1$. It is unknown whether there exist odd perfect numbers. The American number theorist **Carl Pomerance** (1944) showed in 1972 in his dissertation that any odd perfect number has at least seven distinct prime factors. As far as the even perfect numbers is concerned, the problem remains which of the numbers $2^p - 1$ are prime.

10.35 Definition. A prime number of type $2^p - 1$, where $p \in \mathbb{N}$ is called a *Mersenne prime*.

If p is not prime, neither is $2^p - 1$:

10.36 Proposition. *Let $p \in \mathbb{N}^+$ such that $2^p - 1$ is a prime number. Then p is a prime number.*

Proof. From $x^n - 1 = (x-1)\sum_{k=0}^{n-1} x^k$ for $x, n \in \mathbb{N}^+$ it follows that $x - 1 \mid x^n - 1$. So if $x^n - 1$ is a prime number, then $x - 1 = 1$ or $n = 1$. Suppose $p = ab$ with $a, b \in \mathbb{N}^+$. Since $(2^a)^b - 1$ is a prime number it follows that $2^a - 1 = 1$ or $b = 1$. If $2^a - 1 = 1$, then $a = 1$. So p is a prime number. □

The converse is not true: $2^{11} - 1 = 23 \cdot 89$. On the site `www.mersenne.org` a table of all known Mersenne primes is given. Now (September 9, 2024) there are 51 Mersenne primes known. The last one dates from December 21st 2018. Because there is for this kind of numbers an efficient prime test of **Lucas**, improved by the American number theorist **D.H. Lehmer** (1905-1991), the greatest known Mersenne prime usually is the greatest known prime number. The 51-st known Mersenne prime by now is also the greatest ever found:

$$2^{82589933} - 1.$$

In the decimal notation it has 24862048 digits. It is unknown whether a smaller one exists that has not been found.

10.5.2 Euler's totient function

10.37 Definition. Let $n \in \mathbb{N}^+$. Its *totient* $\varphi(n)$ is the number of natural numbers $a < n$ satisfying $\gcd(a, n) = 1$, so

$$\varphi(n) = \#\{\, a \in \mathbb{N}_n \mid \gcd(a, n) = 1 \,\}.$$

The function $\varphi\colon n \mapsto \varphi(n)$ is called the *totient function* (or *Euler's totient function*).

Thus we have an arithmetic function φ with $\varphi(1) = 1$, $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(4) = 2$, $\varphi(5) = 4$, $\varphi(6) = 2$. Clearly $\varphi(n) < n$ if $n > 1$. We also have:

10.38 Proposition. *Let $n \in \mathbb{N}^+$. Then:*

$$n \text{ is a prime number} \iff \varphi(n) = n - 1.$$

Proof.

$\Rightarrow$: Let n be a prime number. Then for all $a \in \mathbb{N}^+$ we have $n \nmid a \iff \gcd(a, n) = 1$. So $\varphi(n) = n - 1$.

$\Leftarrow$: Since $\varphi(n) = n - 1$ (and so $n \neq 1$) we have $\gcd(a, n) = 1$ for all a with $1 \leq a < n$. So there is no divisor d of n with $1 < d < n$, that is n is a prime number. □

For prime powers the totient is easy to determine:

10.39 Proposition. *Let p be a prime number and $k \in \mathbb{N}^+$. Then $\varphi(p^k) = p^k - p^{k-1}$.*

Proof. From $\gcd(a, p^k) \mid p^k$ it follows that $\gcd(a, p^k) = 1 \iff p \nmid a$. There are p^{k-1} multiples of p in $\mathbb{N}_{p^k}$, namely the numbers $0, p, 2p, \ldots, (p^{k-1} - 1)p$. □

10.40 Proposition. *Let $n \in \mathbb{N}^+$. Then $\sum_{d|n} \varphi(d) = n$.*

Proof. Consider the n rational numbers

$$\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}.$$

These are the rational numbers r with $0 \leq r < 1$ and n as a denominator. The least denominator of $\frac{a}{n}$ is $\frac{n}{\gcd(a,n)}$. For each divisor d of n among these rational numbers there are $\varphi(d)$ having d as least denominator. Thus we obtain a partition of $\{\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\}$ where the classes are formed by rational numbers having the same least denominator. From this the proposition follows. □

10.41 Corollary. *The Euler totient function is a multiplicative arithmetic function.*

In chapter 13 another and maybe more conceptual proof of this will be given.

Proof. Proposition 10.40 states that $\varphi * 1 = \mathrm{id}$. Because id is multiplicative, the function φ is by Corollary 10.31 multiplicative as well. □

10.42 Theorem. *Let $n \in \mathbb{N}^+$. Then*

$$\varphi(n) = n \prod_{p|n} \Big(1 - \frac{1}{p}\Big).$$

Proof. This easily follows from Corollary 10.41 and proposition 10.39:

$$\begin{aligned} \varphi(n) &= \prod_{p} \varphi(p^{v_p(n)}) = \prod_{p|n} (p^{v_p(n)} - p^{v_p(n)-1}) \\ &= \prod_{p|n} p^{v_p(n)} \cdot \prod_{p|n} \Big(1 - \frac{1}{p}\Big) = n \cdot \prod_{p|n} \Big(1 - \frac{1}{p}\Big). \end{aligned}$$ □

10.43 Example. We compute the totient of 1000. Because 2 and 5 are the only prime divisors of 1000 we have:

$$\varphi(1000) = 1000 \cdot (1 - \tfrac{1}{2})(1 - \tfrac{1}{5}) = 1000 \cdot \tfrac{1}{2} \cdot \tfrac{4}{5} = 400.$$

Corollary 10.41 and proposition 10.39 can also be used directly:

$$\varphi(1000) = \varphi(2^3)\varphi(5^3) = (8 - 4)(125 - 25) = 4 \cdot 100 = 400.$$

Exercises

1. Let r be a rational number $\neq 0$. Show that $r \in \mathbb{Z}$ if and only if $v_p(r) \geq 0$ for all prime numbers p.
2. Let r be a rational number and let $n \in \mathbb{N}^+$. Show that $r^n \in \mathbb{Z}$ implies $r \in \mathbb{Z}$.
3. Show that there is no rational number x satisfying $x^3 = x + 1$.
4. Which natural numbers $\neq 0$ have exactly 2 positive divisors? Which 3? And which 4?
5. Compute $\tau(252525)$, $\sigma(252525)$ and $\varphi(252525)$.
6. Show that for all $n \in \mathbb{N}^+$:
$$\varphi(2n) = \varphi(n) \iff n \text{ is odd.}$$
7. Let p be a prime number and let the p-adic notation of the natural number $n \geq 1$ be as follows:
$$n = [c_{r-1}, \ldots, c_0]_p.$$

(i) Prove that $p^k \mid n \iff c_i = 0$ for all $i < k$.

(ii) Prove that $v_p(n) = k \iff c_k \neq 0$ and $c_i = 0$ for all $i < k$.

8. Let p be a prime number and let $n \in \mathbb{N}^+$.

(i) Show that for $k \in \mathbb{N}$

$$\#(\{\, a \in \underline{n} \mid v_p(a) = k \,\}) = \left\lfloor \frac{n}{p^k} \right\rfloor - \left\lfloor \frac{n}{p^{k+1}} \right\rfloor.$$

(ii) Show that

$$v_p(n!) = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots.$$

(There are only finitely many nonzero terms.)

(iii) Let $[c_{r-1}, \ldots, c_0]_p$ be the p-adic notation of n. Prove that

$$v_p(n!) = [c_{r-1}, \ldots, c_1]_p + [c_{r-1}, \ldots, c_2]_p + [c_{r-1}, \ldots, c_3]_p + \cdots + [c_{r-1}]_p.$$

(iv) With how many zeros does the decimal notation of 70! end? And in the hexadecimal notation?

9. (i) Show that the arithmetic function $\sigma_2 \colon n \mapsto \sum_{d|n} d^2$ is multiplicative.

(ii) Determine a formula for $\sigma_2(n)$.

10. Let $\rho(n)$ be the number of prime factors of n:

$$\rho(n) = \sum_p v_p(n).$$

The arithmetic function λ is defined by $\lambda(n) = (-1)^{\rho(n)}$. (The function λ is sometimes called the *Liouville function*.)

(i) Show that λ is strictly multiplicative.

(ii) Prove that

$$\sum_{d|n} \lambda(d) = \begin{cases} 1 & \text{if } n \text{ is a square,} \\ 0 & \text{if } n \text{ is not a square.} \end{cases}$$

11. Let $n \in \mathbb{N}^+$. Show that the number of $(k, l) \in {\mathbb{N}^+}^2$ with $\operatorname{lcm}(k, l) = n$ equals $\tau(n^2)$.

12. At Amazon a poster could be ordered displaying the 51-st Mersenne prime number. That number is given in its decimal representation. A magnifier is included so as to make the more than 25 million digits visible. What would it look like in its binary representation? And hexadecimal?

13. Show that for all $n \in \mathbb{N}^+$ the number $\tau(n)$ is odd if and only if n is a square.

14. For which $n \in \mathbb{N}^+$ is $\sigma(n)$ odd?

15. Show that $\prod_{d|n} d = n^{\frac{\tau(n)}{2}}$. (Remark: if $\tau(n)$ is odd, then n is a square and the formula is to be read as $\prod_{d|n} d = \sqrt{n}^{\,\tau(n)}$.)

16. The arithmetic function χ is defined by $\chi(n) = \sum_{d|n} \frac{1}{d}$.
    (i) Show that χ is multiplicative.
    (ii) Prove that $\chi(n)\varphi(n) = \prod_{p|n}(p^{v_p(n)} - \frac{1}{p})$ for all $n \in \mathbb{N}^+$.

17. Prove that for all $m, n \in \mathbb{N}^+$:
    (i) $\varphi(m)\varphi(n) = \varphi(\gcd(m,n))\varphi(\operatorname{lcm}(m,n))$,
    (ii) $\varphi(m)\varphi(\gcd(m,n)) = \gcd(m,n)\varphi(m)\varphi(n)$.

18. Let n be a natural number ≥ 2. Show that the sum of all $k \in \underline{n}$ equals $\frac{1}{2}n\varphi(n)$.

19. (i) Let $n \in \mathbb{Z}$. Determine (depending on n)
$$\gcd(n^2 + n, 41) \quad \text{and} \quad \gcd(41n^2, n+1).$$
    (ii) Show that for infinitely many $n \in \mathbb{N}$ none of the numbers
$$n, \quad n^2 + n + 41 \quad \text{and} \quad 41n^2 + n + 1$$
    is prime.

20. (i) For $x, y \in \mathbb{N}^+$ with $\gcd(x,y) = 1$ let $4xy$ be a square of a natural number. Show that both x and y are squares of a natural number.
    (ii) For $x, y \in \mathbb{N}^+$ with $\gcd(x,y) = 1$ let $2xy$ be a square of a natural number. Show that exactly one of the numbers x and y is a square of a natural number.

21. (i) The arithmetic function f is defined by $f(n) = \mu(n)n$. Show that f is multiplicative.
    (ii) Show that f satisfies $f * \mathrm{id} = \mathbf{1}$.

22. An arithmetic function g is called invertible if there exists an arithmetic function h such that $g * h = \mathbf{1}$. Prove:
$$g \text{ is invertible} \iff g(1) \neq 0.$$

23. An alternative proof of theorem 10.17. Let (x, y, z) be a primitive Pythagorean triple with y even. Put $d = \gcd(x+z, y)$, $u = \frac{x+z}{d}$ and $v = \frac{y}{d}$.
    (i) Show that $\dfrac{2uv}{u^2+v^2} = \dfrac{y}{z}$ and $\dfrac{u^2-v^2}{u^2+v^2} = \dfrac{x}{z}$.
    (ii) Show that $\gcd(u,v) = 1$ and that from $\dfrac{2uv}{u^2+v^2} = \dfrac{y}{z}$ it follows that u and v are not both odd.
    (iii) Prove that $\gcd(u^2 - v^2, u^2 + v^2) = 1$.
    (iv) Prove that $x = u^2 - v^2$, $z = u^2 + v^2$ and $y = 2uv$.

24. Let the natural number a be odd and let $\frac{a^2-1}{8}$ be a prime number. Determine a.

25. (i) Show that there is no integer x such that $(x^2 - 18)x = 9$.
    (ii) Show that there is no rational number x such that $x^3 = 18x + 9$.

26. Let N be a natural number having 3 as the last digit in its decimal representation. Show that there are no natural numbers c and n such that

$$\frac{1}{N} = \frac{c}{10^n}.$$

27. (i) Prove that for all $n \in \mathbb{N}^+$ there is a $k \in \mathbb{N}$ such that $\varphi^k(n) = 1$.
(ii) Prove that for all $k \in \mathbb{N}$ there is an $n \in \mathbb{N}^+$ such that $\varphi^k(n) \neq 1$.

11 Combinatorics

Combinatorics is the art of counting, especially for obtaining results about finite structures such as graphs. In chapter 1 for example we considered the number of moves needed for the solution of the Tower of Hanoi puzzle using simple arithmetic.

Binomial coefficients are numbers of subsets of a given size. In mathematics these numbers are often used: in counting problems, as in section 11.3, and also in an algebraic context, for example when elaborating expressions like $(a+b)^n$—which in fact explains their name. Polynomial sequences are sequences in which the n-th term is given by a polynomial in n, for example the sequence of squares: the n-th term then is n^2. In section 11.4 we study partial sum sequences and difference sequences of polynomial sequences.

Stirling numbers of the second kind are numbers of partitions of a given size. For these numbers there are formulas too, formulas which are somewhat more complicated than those for binomial coefficients. Stirling numbers of the first kind are treated in the next chapter.

11.1 Injective Maps and Subsets

Let A and B be finite sets such that $\#(A) \leq \#(B)$. The image of an injective map $f\colon A \to B$ is a subset of B with $\#(A)$ elements. We will derive formulas for the number of injective maps between finite sets and for the number of subsets of a finite set with a given number of elements.

11.1.1 Injective maps

11.1 Notation. Let A and B be sets. We denote the set of injective maps from A to B by $\mathrm{Inj}(A,B)$.

This is not a generally accepted notation. It is only used in this section.

11.2 Proposition. *Let A and B be finite sets such that $\#(A) \leq \#(B)$. Then*

$$\#(\mathrm{Inj}(A,B)) = \frac{\#(B)!}{(\#(B)-\#(A))!}.$$

Proof. We prove the formula by mathematical induction on $\#(A)$. If $\#(A) = 0$, then A is empty and so there is just one injective map from A to B.

Suppose the proposition is true for all sets A with $\#(A) = k$. Let A be a set of $k+1$ elements and B arbitrary, say a set of n elements, where $n \geq k+1$. Since $k+1 > 0$ and so $A \neq \emptyset$, we can fix an $a \in A$. By assigning the image of this a to an injective map we obtain a surjection

$$\mathrm{Inj}(A, B) \to B,\ f \mapsto f(a).$$

This map induces a partition of $\mathrm{Inj}(A, B)$ with for each $b \in B$ a class

$$\{\, f \in \mathrm{Inj}(A, B) \mid f(a) = b \,\}.$$

This class has as many elements as $\mathrm{Inj}(A \setminus \{a\}, B \setminus \{b\})$, and by the induction hypothesis this number is $(n-1)(n-2)\cdots(n-k)$. There are n of these classes, all of the same size. So the total number is $n(n-1)(n-2)\cdots(n-k)$. □

The same proof, but less formal. An injective map from A to B is made by subsequently choosing images for the elements of A. Assume that the elements are numbered from 1 to k. For the first element there are n possible images, then for the second one $n-1$, etc. So in total $n(n-1)(n-2)\cdots(n-k+1)$.

A special case:

11.3 Corollary. *Let A and B be finite sets such that $\#(A) = \#(B) = n$. Then there are $n!$ bijective maps from A to B.*

Proof. By theorem 5.37 injective maps from A to B are bijective. □

11.4 The birthday paradox. What is the probability that in a company of N persons two of them have the same birthday? Let's ignore February 29th. The probability that all persons have *different* birthdays equals

$$\frac{\text{the number of injective maps } \underline{N} \to \underline{365}}{\text{the number of maps } \underline{N} \to \underline{365}}.$$

So this probability is $\frac{365 \cdot 364 \cdots (365-N+1)}{365^N} = \prod_{j=0}^{N-1}(1 - \frac{j}{365})$. By increasing N the probability decreases: for $N = 1$ it is 1, for $N > 365$ it is 0. From which N onwards is it less then $\frac{1}{2}$? This question is known as the *birthday problem*. The solution is $N = 23$, which is less then most people would guess. That is why it is often called the birthday paradox. (Generally, when dealing with injective maps to $\underline{M}$, this N is in the order of magnitude of $\sqrt{M}$.)

Python

The code for the computation of factorials is simple.

combinatorics.py

```
def factorial(n):
    i, fact = 0, 1
    while i < n:
        i, fact = i + 1, fact * (n - i)
    return fact
```

```
>>> factorial(77)
145183092028285869634070784086308284983740379224208358846781574688061
991349156420080065207861248000000000000000000
```

11.1.2 Subsets with k elements

11.5 Notation. Let A be a finite set with n elements and let k be a natural number such that $k \leq n$. We denote by $\mathcal{P}_k(A)$ the set consisting of all subsets of A with k elements, so:

$$\mathcal{P}_k(A) = \{\, U \in \mathcal{P}(A) \mid \#(U) = k \,\}.$$

We define:

11.6 Definition. Let n and k be natural numbers with $k \leq n$. Then

$$\binom{n}{k} = \#(\mathcal{P}_k(\underline{n})).$$

This number is called a *binomial coefficient*. (In the next section it will become clear why it is called this way).

> In the definition we took $A = \underline{n}$. Since, for $\#(A) = n$ it is not hard to see that $\mathcal{P}_k(A) \approx \mathcal{P}_k(\underline{n})$, any set A with n elements could have been taken.

Let A be a finite set with $\#(A) = n$. Let k be a natural number such that $0 < k < n$. (So $n \geq 2$.) We fix an element a of A. Then two kinds of subsets of A with k elements can be distinguished:

a) subsets U such that $a \in U$,
b) subsets U such that $a \notin U$.

Subsets of the first kind correspond to subsets of $A \setminus \{a\}$ with $k-1$ elements. There are $\binom{n-1}{k-1}$ of these. Subsets of the second kind correspond to subsets of $A \setminus \{a\}$ with k elements and of these there are $\binom{n-1}{k}$. So we derived:

$$\binom{0}{0}$$
$$\binom{1}{0}\ \binom{1}{1}$$
$$\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2}$$
$$\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3}$$
$$\binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4}$$
$$\binom{5}{0}\ \binom{5}{1}\ \binom{5}{2}\ \binom{5}{3}\ \binom{5}{4}\ \binom{5}{5}$$
$$\binom{6}{0}\ \binom{6}{1}\ \binom{6}{2}\ \binom{6}{3}\ \binom{6}{4}\ \binom{6}{5}\ \binom{6}{6}$$
$$\binom{7}{0}\ \binom{7}{1}\ \binom{7}{2}\ \binom{7}{3}\ \binom{7}{4}\ \binom{7}{5}\ \binom{7}{6}\ \binom{7}{7}$$

Figure 11.1: Pascal's triangle

11.7 Proposition. *Let n and k be natural numbers with $0 < k < n$. Then*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$ □

Other, more obvious, properties are:

11.8 Proposition. *Let n and k be natural numbers with $k \leq n$. Then:*

$$\binom{n}{0} = \binom{n}{n} = 1 \qquad \textit{and} \qquad \binom{n}{m} = \binom{n}{n-m}.$$

Proof. This follows from:

a) The empty subset of $\underline{n}$ is the only subset of 0 elements.
b) $\underline{n}$ is the only subset of $\underline{n}$ of n elements.
c) Subsets U of $\underline{n}$ with k elements correspond to subsets V of $\underline{n}$ with $n-k$ elements: take $V = \underline{n} \setminus U$. In other words: the map $\mathcal{P}_k(\underline{n}) \to \mathcal{P}_{n-k}(\underline{n})$, $U \mapsto \underline{n} \setminus U$ is bijective. □

The binomial coefficients can nicely be displayed in *Pascal's triangle*, a triangular diagram, see Figure 11.1. Using the propositions 11.7 and 11.8 the triangle is easily computed, see Figure 11.2. It can be seen as an inductive definition of the sequence of rows of the triangle. The numbers $\binom{n}{k}$ in Figure 11.1 can be seen as the number of paths in the directed graph of Figure 11.3 from the top downwards. The number of paths ending in (n, k) equals the sum of the numbers of paths ending in the vertices immediately above (n, k).

A path is formed by repeated choices for left or right: descending one vertex there is a choice between the left and the right vertex. The row you end at depends on

Blaise Pascal (Clermont 1623 – Pasequences 1662)

Pascal was a French mathematician who already as a youth was occupied with mechanical addition of numbers. The machines he built are in a sense the precursors of the computer. Pascal's triangle is named after him, but was then already known for several ages. Pascal did use the triangle in connection with probability calculus and he also has derived various properties of the triangle. He had a clear way of reasoning, though he preferred to use words over formulas. He also contributed to projective geometry.

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}
$$

Figure 11.2: Pascal's triangle computed

the number of choices made. After n choices you end in row n. When of a total of n choices k times has been chosen for right, you end in the vertex (n, k). Thus the paths ending in this vertex correspond to subsets of $\{1, 2, \ldots, n\}$ with k elements.

In Pascal's triangle a lot of remarkable properties can be discovered. For instance that the sum of the numbers in row n equals 2^n. This is the total number of subsets of a set which has n elements. In that row these numbers are split up according to the number of elements of the subsets.

Algorithm

The binomial coefficients are determined by propositions 11.7 and 11.8 and these can be used for their computation. That is the way Pascal's triangle in Figure 11.2 was computed. When designing an algorithm it is important that the same numbers are not computed over and over again. For the computation of a binomial coefficient only the numbers above that number in Pascal's triangle are relevant. So there is a

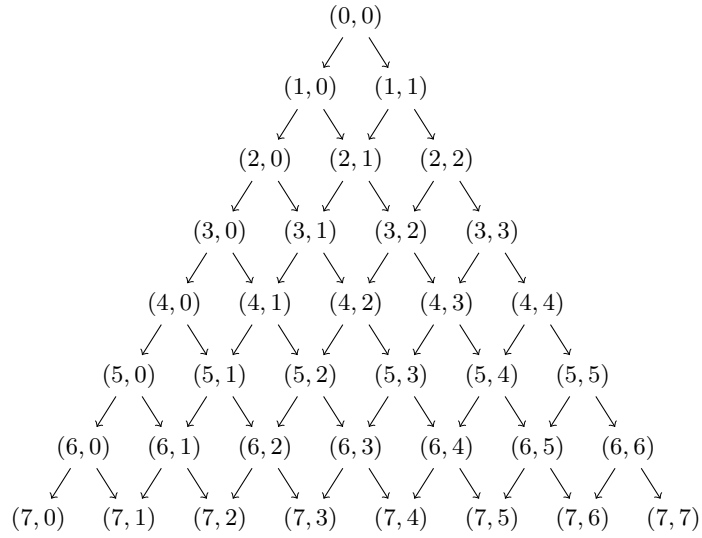


Figure 11.3: Directed graph with vertices (n, k)

'parallelogram' of numbers which have to be computed. The list of numbers along the side left above consists of ones only. The list of numbers directly below them starts with a 1 in top and can be computed using the first list. Thus every next list can be computed and finally the binomial coefficient to be computed is the last number in the last list. See Figure 11.4 for the computation of $\binom{7}{3}$.

Python

combinatorics.py

```
def comb(n, k):
    list0 = (n - k + 1) * [1]
    i = j = 0
    while i < k:
        list1 = [1]
        while j < n - k:
            j = j + 1
            list1.append(list1[-1] + list0[j])
        list0 = list1
        i, j = i + 1, 0
    return list0[-1]
```

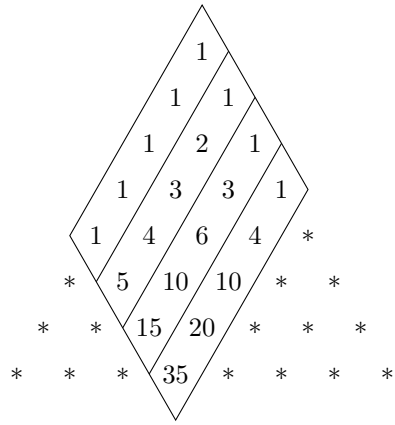


Figure 11.4: The computation of $\binom{7}{3}$

```
>>> comb(1209, 476)
230116158359325490132982607650290019977276112005917688014564482394933
407667069406658633828945204464182765630512325481132705106176377309772
405347735054365224940922421594166068837010493521907843522812670490559
054454495705418953437646984979117129356684864911536999791504325503341
708279703860428588289372343275707934120024223127888167649186881639650
286720
```

So far we computed binomial coefficients using Pascal's triangle. There is also a direct formula for a binomial coefficient, which is often taken as its definition:

11.9 Proposition. *Let k and n be natural numbers such that $k \leq n$. Then*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proof. Let A and B be finite sets with $\#(A) = k$ and $\#(B) = n$. The surjective map

$$\mathrm{Inj}(A, B) \to \mathcal{P}_k(B),\ f \mapsto f_*(A)$$

induces a partition of $\mathrm{Inj}(A, B)$. For every $U \subseteq B$ with $\#(U) = k$ there is a class

$$\{\, f \in \mathrm{Inj}(A, B) \mid f_*(A) = U \,\}.$$

This class has as many elements as there are bijections from A to U. This number is by Corollary 11.3 equal to $k!$. So

$$\#(\mathrm{Inj}(A, B)) = \sum_{U \in \mathcal{P}_k(B)} k! = k! \sum_{U \in \mathcal{P}_k(B)} 1 = k! \cdot \binom{n}{k}.$$

From this it follows that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$ □

This formula can also be derived by verifying the relations which determine the binomial coefficients, or, what amounts to the same, the relations which determine Pascal's triangle.

11.2 Products of Binomials

Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be numbers. We are going to expand the expression

$$(a_1 + b_1)(a_2 + b_2) \cdots (a_n + b_n).$$

The result will be a sum of 2^n terms, each term being a product of n factors. After that we will have a look at special cases: for example all a_i equal, or all a_i equal and the same for all b_j.

11.10 Notation. We will use temporarily the following notation:

$$a_I = \prod_{i \in I} a_i,$$

where $a_1, \ldots, a_n$ are numbers and I a subset of $\underline{n}$. Thus for example if $n = 7$, then

$$a_{\{2,5,7\}} = a_2 a_5 a_7, \quad a_{\{3\}} = a_3 \quad \text{and} \quad a_\emptyset = 1.$$

With this notation some formulas will be more readable, e.g. the formula in the following theorem.

11.11 Theorem. *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be numbers. Then*

$$\prod_{i=1}^{n} (a_i + b_i) = \sum_{I \subseteq \underline{n}} a_I b_{\underline{n} \setminus I},$$

where $\sum_{I \subseteq \underline{n}}$ stands for $\sum_{I \in \mathcal{P}(\underline{n})}$.

Proof. The proof will be by mathematical induction on n. For $n = 0$ on the left hand side we have an empty product and on the right hand side $a_\emptyset b_\emptyset = 1$.

Suppose the formula is correct for some $n \in \mathbb{N}$. Let $a_1, \ldots, a_{n+1}$ and $b_1, \ldots, b_{n+1}$ be numbers. Then indeed

$$\prod_{i=1}^{n+1} (a_i + b_i) = (a_{n+1} + b_{n+1}) \prod_{i=1}^{n} (a_i + b_i) = (a_{n+1} + b_{n+1}) \sum_{I \subseteq \underline{n}} a_I b_{\underline{n} \setminus I}$$

$$= \sum_{I\subseteq\underline{n}} a_I a_{n+1} b_{\underline{n}\setminus I} + \sum_{I\subseteq\underline{n}} a_I b_{\underline{n}\setminus I} b_{n+1} = \sum_{\substack{J\subseteq\underline{n+1}\\ n+1\in J}} a_J b_{\underline{n+1}\setminus J} + \sum_{\substack{J\subseteq\underline{n+1}\\ n+1\notin J}} a_J b_{\underline{n+1}\setminus J}$$

$$= \sum_{J\subseteq\underline{n+1}} a_J b_{\underline{n+1}\setminus J}. \qquad \square$$

Now we consider the special case $a_1 = a_2 = \cdots = a_n = a$.

11.12 Theorem. *Let a and $b_1, \ldots, b_n$ be numbers. Then*

$$\prod_{i=1}^{n}(a + b_i) = \sum_{k=0}^{n} s_k a^{n-k},$$

where $s_k = \sum_{\#(I)=k} b_I$.

Proof. We apply theorem 11.11. Since $a_I = a^{\#(I)}$ we have

$$\prod_{i=1}^{n}(a + b_i) = \sum_{k=0}^{n} \sum_{\#(I)=k} a^k b_{\underline{n}\setminus I} = \sum_{k=0}^{n} \Big(\sum_{\#(I)=k} b_{\underline{n}\setminus I} \Big) a^k$$

$$= \sum_{k=0}^{n} s_{n-k} a^k = \sum_{k=0}^{n} s_k a^{n-k}. \qquad \square$$

This formula often occurs in the following form:

$$\prod_{i=1}^{n}(x - a_i) = \sum_{k=0}^{n}(-1)^k s_k x^{n-k},$$

where $s_k = \sum_{\#(I)=k} a_I$. The right hand side is a polynomial in x of degree n and the left hand side is a product of n factors each of the form $x - a$. Equating the polynomial to 0 gives an equation of degree n with $x = a_1, \ldots, x = a_n$ as solutions. Starting from the solutions you find an equation. When solving an equation somehow one has to go the other way round.

For $n = 4$ we have for example:

$$\begin{aligned}
s_0 &= 1,\\
s_1 &= a_1 + a_2 + a_3 + a_4,\\
s_2 &= a_1a_2 + a_1a_3 + a_1a_4 + a_2a_3 + a_2a_4 + a_3a_4,\\
s_3 &= a_1a_2a_3 + a_1a_2a_4 + a_1a_3a_4 + a_2a_3a_4,\\
s_4 &= a_1a_2a_3a_4.
\end{aligned}$$

Isaac Newton (Woolsthorpe 1642 – London 1727)

Newton was both a mathematician and a physicist. He invented what is now called calculus. So did the German mathematician **Gottfried Leibniz** (1646–1716), who also developed its present day notation. Calculus is the mathematical foundation of the Newtonian mechanics, which for example explains the shape of the orbits of the planets.

The Binomial Theorem

The numbers $\binom{n}{k}$ are called binomial coefficients because they occur as coefficients when expanding a power of a binomial.

11.13 Binomial Theorem. *Let a and b be numbers, and let $n \in \mathbb{N}$. Then*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Proof. We apply theorem 11.11 with $b_1 = \cdots = b_n = b$. Since $b_I = b^{\#(I)}$ we now have $s_k = \sum_{\#(I)=k} b^k = b^k \sum_{\#(I)=k} 1 = b^k \binom{n}{k}$ and so

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k. \qquad \square$$

The Binomial Theorem is often named after Newton, but the theorem was known as early as five centuries before Newton. Newton used his calculus to generalize the theorem to exponents in $\mathbb{R}$ (the real numbers) and not just in $\mathbb{N}$:

$$(1+x)^a = \sum_{n=0}^{\infty} \binom{a}{n} x^n$$

for all $x \in (-1, 1)$, where $\binom{a}{n}$ is defined as will be done in definition 11.27.

We already noticed that the sum of the numbers in row n of Pascal's triangle equals 2^n. Another way to see this is by applying the Binomial Theorem:

$$2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k}.$$

Note that $0^0 = 1$. In the proof we used for example $b^0 = 1$ without saying. If we want the formula to be generally valid, so also for $b = 0$, then it is desirable to agree on $0^0 = 1$, as in fact we did in chapter 4.

11.3 Catalan Numbers

A *stack* is an important data structure in computer science. Data can be stored in a stack with the restriction that adding and deleting data is only possible on the *top* of the stack. We will put the numbers 1 up to n in that order on the stack and will count the number of ways they can leave the stack. We describe this with discs numbered 1 up to n. In Figure 11.5 they are placed on peg 1.

The discs will be placed on the stack (peg 2) by moving them one by one from peg 1 to peg 2. If there are discs on peg 2, then the top disc may be moved to peg 3. So on peg 3 the pins will be placed from bottom to top in the order they leave peg 2. How many of these orders can occur this way?

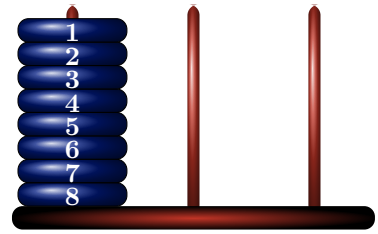


Figure 11.5: Initial position with 8 discs

A final position is reached after exactly $2n$ moves. For the notation of intermediate and final positions the 'history' that led to them can be used. There are two types of moves:

a) type `0`, where a disc is moved from peg 1 to peg 2; this corresponds to the placement of a number on the stack,
b) type `1`, where a disc is moved from peg 2 to peg 3; this corresponds to removal of a number from the stack.

A word of zeros and ones describes which moves have been made and in which order. The word `001011010` indicates that first two times a disc is moved from peg 1 to peg 2, then one from 2 to 3, one from 1 to 2, two from 2 to 3, one from 1 to 2, one from 2 to 3 and finally one from 1 to 2. The number of zeros indicates how many numbers have been placed on the stack and the number of ones how many have left the stack. The word `001011010` describes the position in Figure 11.6. The number of zeros minus the number of ones equals the number of discs on peg 2. For any position the number of ones does not exceed the number of zeros. So a word of ones and zeros describes a position if and only if in every initial segment the number of ones does not exceed the number of zeros. Such a word will be called an *admitted* word. The given word is admitted, since its initial segments are

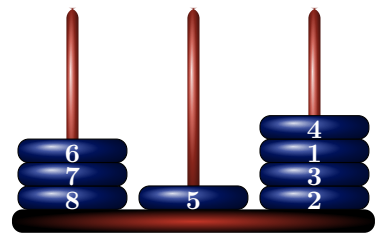


Figure 11.6: The position 001011010

`0, 00, 001, 0010, 00101, 001011, 0010110, 00101101, 001011010`

Eugène Charles Catalan (Bruges 1814 – Liège 1894)

Catalan is mainly known for the numbers named after him. He studied the number of ways parentheses can be placed in a word of given length. A product abc can be seen as $(ab)c$ and as $a(bc)$. In case of 4 factors there are 5 ways: $(a(bc))d$, $((ab)c)d$, $(ab)(cd)$, $a((bc)d)$ and $a(b(cd))$. In case of a product of $n+1$ factors it can be done in c_n ways, see exercise 11.

and for each of these words the number of ones does not exceed the number of zeros. The number of final positions equals the number of admitted words which contain n zeros and n ones. The problem of the number of possible orders on peg 3 is thus translated into the question: how many of these words are there?

The binomial coefficient $\binom{n}{k}$ is equal to the number of paths in Figure 11.3 from $(0,0)$ to (n,k). Admitted words correspond to paths through vertices (i,j) with $j \leq i-j$, see Figure 11.7.

For the number of admitted words we introduce a notation:

11.14 Definition and notation. Let n and k be natural numbers with $k \leq \lfloor \frac{n}{2} \rfloor$. Then $\left\langle {n \atop k} \right\rangle$ denotes the number of admitted words in $\{0,1\}$ of length n with exactly k ones. The *n-th Catalan number* c_n is the number $\left\langle {2n \atop n} \right\rangle$.

Catalan numbers can be computed in the same way as binomial coefficients, see Figure 11.8.

The numbers $\left\langle {n \atop k} \right\rangle$ are determined by:

$$\begin{cases} \left\langle {n \atop 0} \right\rangle = 1 & \text{for all } n \in \mathbb{N}, \\ \left\langle {2n \atop n} \right\rangle = \left\langle {2n-1 \atop n-1} \right\rangle & \text{for all } n \in \mathbb{N}^+, \\ \left\langle {n \atop k} \right\rangle = \left\langle {n-1 \atop k-1} \right\rangle + \left\langle {n-1 \atop k} \right\rangle & \text{for all } k,n \in \mathbb{N} \text{ with } 0 < k < \lfloor \frac{n}{2} \rfloor, \end{cases}$$

that is in every vertex the number equals the sum of the numbers in the vertices directly above in the scheme of Figure 11.8. Starting with a 1 in top, the whole scheme can be completed row after row from top to bottom. After having computed this way some of the numbers $\left\langle {n \atop k} \right\rangle$ the following theorem emerges:

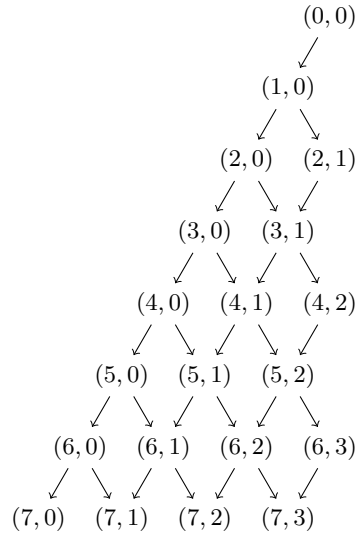


Figure 11.7: Directed graph with vertices (n, k) where $k \leq n - k$

11.15 Theorem. *Let n and k be natural numbers such that $0 < k \leq \lfloor \frac{n}{2} \rfloor$. Then*

$$\left\langle {n \atop k} \right\rangle = \binom{n}{k} - \binom{n}{k-1}.$$

Proof. Put

$$c(n,k) = \begin{cases} \binom{n}{k} - \binom{n}{k-1} & \text{if } 0 < k \leq \lfloor \frac{n}{2} \rfloor, \\ 1 & \text{if } k = 0. \end{cases}$$

and show that these numbers satisfy the rules that determine the Catalan numbers:

a) $c(n, 0) = 1$ by definition.
b) We have $c(2, 1) = \binom{2}{1} - \binom{2}{0} = 2 - 1 = 1 = c(1, 0)$. And for $n \geq 2$

$$\begin{aligned} c(2n, n) &= \binom{2n}{n} - \binom{2n}{n-1} \\ &= \binom{2n-1}{n-1} + \binom{2n-1}{n} - \binom{2n-1}{n-2} - \binom{2n-1}{n-1} \\ &= \binom{2n-1}{n-1} - \binom{2n-1}{n-2} + \binom{2n-1}{n} - \binom{2n-1}{n-1} \\ &= c(2n-1, n-1). \end{aligned}$$

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & & & \\
 & 1 & & 3 & & 2 & & & \\
1 & & 4 & & 5 & & & & \\
 & 1 & & 5 & & 9 & & 5 & \\
1 & & 6 & & 14 & & 14 & &
\end{array}
$$

Figure 11.8: Computation of the numbers $\left\langle {n \atop k} \right\rangle$

c) For $n, k \in \mathbb{N}$ such that $1 < k < \lfloor \frac{n}{2} \rfloor$:

$$
\begin{aligned}
c(n-1,k-1) + c(n-1,k) &= \binom{n-1}{k-1} - \binom{n-1}{k-2} + \binom{n-1}{k} - \binom{n-1}{k-1} \\
&= \binom{n-1}{k-1} + \binom{n-1}{k} - \binom{n-1}{k-2} - \binom{n-1}{k-1} = \binom{n}{k} - \binom{n}{k-1} = c(n,k)
\end{aligned}
$$

and for $k = 1$:

$$
c(n-1,0) + c(n-1,1) = 1 + \binom{n-1}{1} - \binom{n-1}{0} = n - 1 = c(n,1). \quad \square
$$

In particular we now have a formula for the n-th Catalan number:

11.16 Theorem. *Let $n \in \mathbb{N}$. Then*

$$
c_n = \frac{1}{n+1}\binom{2n}{n}.
$$

Proof. From theorem 11.15 follows

$$
\begin{aligned}
c_n = \left\langle {2n \atop n} \right\rangle &= \binom{2n}{n} - \binom{2n}{n-1} = \binom{2n}{n} - \frac{(2n)!}{(n-1)!(n+1)!} \\
&= \binom{2n}{n} - \frac{n}{n+1}\binom{2n}{n} = \frac{1}{n+1}\binom{2n}{n}.
\end{aligned}
\qquad \square
$$

So the number of possible orders of the discs on peg 3 in case of n discs equals $\frac{1}{n+1}\binom{2n}{n}$.

A Direct Way

The number $\left\langle {n \atop k} \right\rangle$ of admitted words of zeros and ones of length n with k ones can also be computed directly by counting the number of words that are *not* admitted. That number has to be $\binom{n}{k-1}$. How to see this?

Let $c_1c_2c_3\ldots c_n$ be a not admitted word with k ones, where $0 < k \leq \lfloor \frac{n}{2} \rfloor$. Then there are numbers m with $0 < m \leq n$ such that $c_1c_2\ldots c_m$ is a not admitted word (for example $m = n$). Now let m with $0 < m \leq n$ be the least of these numbers. Then the initial segment of length m ends with a `1` and the number of ones exceeds the number of zeros by 1. For example

`0100111110000000001011111100000`

is a not admitted word of length 31 and the least not admitted initial segment is

`0100111.`

Now replace in the smallest not admitted initial segment all zeros by ones and all ones by zeros, leaving the remainder unchanged. In the example this gives the word

`1011000110000000001011111100000.`

The word thus obtained is a word of length n with $k-1$ ones. There are $\binom{n}{k-1}$ of such words. Each of these words is obtained this way from a not admitted word of length n with k ones, as is seen as follows. A word of length n with $k-1$ ones has less ones than zeros and there has to be a smallest initial segment with less ones than zeros. Replace in this segment all ones by zeros and all zeros by ones. The resulting word will be not admitted and will have k ones. It is easily seen that this defines a correspondence between not admitted words of length n with k ones and words of the same length with $k-1$ ones. Of these there are $\binom{n}{k-1}$.

A Recursive Description

We have seen that the numbers $\left\langle {n \atop k} \right\rangle$ can be computed in a similar way as was done for the binomial coefficients. For the Catalan numbers $c_n = \left\langle {2n \atop n} \right\rangle$ there is a recursive description which can be used to compute them one by one:

$$\begin{cases} c_0 = 1, \\ c_n = \sum_{k=0}^{n-1} c_k c_{n-1-k} \quad \text{for all } n \in \mathbb{N}^+. \end{cases}$$

This gives:

$$c_0 = 1,$$

$$
\begin{aligned}
c_1 &= 1 \cdot 1 = 1,\\
c_2 &= 1 \cdot 1 + 1 \cdot 1 = 2,\\
c_3 &= 1 \cdot 2 + 1 \cdot 1 + 2 \cdot 1 = 5,\\
c_4 &= 1 \cdot 5 + 1 \cdot 2 + 2 \cdot 1 + 5 \cdot 1 = 14,\\
c_5 &= 1 \cdot 14 + 1 \cdot 5 + 2 \cdot 2 + 5 \cdot 1 + 14 \cdot 1 = 42,\\
&\vdots
\end{aligned}
$$

This recursive description can be understood as follows. Let n be a natural number ≥ 1. To each admitted word of n zeros and n ones there is a least $m > 0$ such that the initial segment of length $2m$ has as many ones as zeros. This initial segment starts with a `0` and ends with a `1`. In between this `0` and `1` there is an admitted word of length $2m - 2$. So the word consists subsequently of

1. a zero,
2. an admitted word of $m - 1$ zeros and $m - 1$ ones,
3. a one,
4. an admitted word of $n - m$ zeros and $n - m$ ones.

For a given m there are $c_{m-1}c_{n-m}$ of such words. So the total number is

$$c_0c_{n-1} + c_1c_{n-2} + c_2c_{n-3} + \cdots + c_{n-2}c_1.$$

11.4 Polynomial Sequences

The n-th term of the sequence (n^2) is a polynomial in n. It is easily shown by induction that $\sum_{k=0}^{n-1} k^2 = \frac{1}{6}n(n-1)(2n-1)$. So the sequence with the sum of the first n squares as the nth term is a polynomial sequence as well. In this section we study such sequences in general.

11.17 Definition. A *polynomial sequence* is a sequence $(f(n))$, where $f(n)$ is a polynomial in n. The *degree* of a polynomial sequence is the degree of the polynomial $f(n)$. (The sequence (0) is the only polynomial sequence with no degree.)

If $f(x)$ is a polynomial, say $f(x) = c_0x^m + c_1x^{m-1} + \cdots + c_{m-1}x + c_m$, then $f(x+1) - f(x)$ is a polynomial as well:

$$f(x+1) - f(x) = c_0((x+1)^m - x^m) + c_1((x+1)^{m-1} - x^{m-1}) + \cdots + c_{m-1}$$

and we have

$$(x+1)^k - x^k = \sum_{l=0}^{k-1} \binom{k}{l} x^l.$$

If $\deg(f) = m > 0$, then the degree of $f(x+1) - f(x)$ equals $m - 1$. The leading coefficient of $f(x+1) - f(x)$ is $c_0 \cdot m$.

11.18 Definitions. Let (a_n) be a sequence of numbers. The sequence (d_n) with $d_n = a_{n+1} - a_n$ is called the *difference sequence* of the sequence (a_n). The sequence (s_n) with $s_n = \sum_{k=0}^{n-1} a_k$ is called the *partial sum sequence* of the sequence (a_n) or also the *series* associated to the sequence (a_n). The s_n are called the *partial sums* of (a_n) and the a_n the *general term* of the series (s_n).

Note that the difference sequence of the partial sum sequence of (a_n) equals the original sequence (a_n). The partial sum sequence of the difference sequence of (a_n) is the sequence $(a_n - a_0)$; the terms differ from the terms of (a_n) by the constant a_0.

11.19 Notation. For a polynomial $f(x)$ we denote the polynomial $f(x+1) - f(x)$ as $(\Delta f)(x)$. The difference sequence of a polynomial sequence $(f(n))$ is then the polynomial sequence $((\Delta f)(n))$.

Δ can be seen as a transformation of the set of polynomials. There is a unique polynomial $f(x)$ with $(\Delta f)(x) = f(x)$, namely $f(x) = 0$, the 0-polynomial.

11.20 Example. The sequence of the cubes of the natural numbers is a polynomial sequence. It is the sequence (n^3). We take the difference sequence, the difference sequence of the difference sequence, and so on:

$$\begin{array}{ccccccccccc}
0 & & 1 & & 8 & & 27 & & 64 & & 125 & & 216 & & 343 & \cdots \\
& 1 & & 7 & & 19 & & 37 & & 61 & & 91 & & 127 & & \cdots \\
& & 6 & & 12 & & 18 & & 24 & & 30 & & 36 & & \cdots & \\
& & & 6 & & 6 & & 6 & & 6 & & 6 & & \cdots & & \\
& & & & 0 & & 0 & & 0 & & 0 & & \cdots & & &
\end{array}$$

We have $(\Delta f)(n) = 3n^2+3n+1$, $(\Delta^2 f)(n) = 6n+6$, $(\Delta^3 f)(n) = 6$ and $(\Delta^k f)(n) = 0$ for all $k > 3$.

11.21 Proposition. *Let $f(x)$ be a polynomial of degree m and let a be the leading coefficient of f. Then $(\Delta^m f)(x) = m! \cdot a$.*

Proof. Because with every application of Δ the degree of the polynomial goes down by 1, the polynomial $(\Delta^m f)(x)$ is of degree 0, that is it is a nonzero constant. Moreover with every application of Δ the leading coefficient is multiplied by the degree. □

So the course of a polynomial $f(x)$ of degree m under the transformation Δ is

$$f(x), (\Delta f)(x), (\Delta^2 f)(x), \ldots, (\Delta^m f)(x), 0, 0, 0, 0, \ldots$$

with $(\Delta^m f)(x)$ a nonzero constant.

The partial sum sequence of a polynomial sequence has this polynomial sequence as its difference sequence. We will show that the partial sum sequence is also a polynomial sequence. In particular this means that the partial sum sequence of (n^m) (where $m \in \mathbb{N}$), the sequence of m-th powers, is a polynomial sequence.

11.22 Notation. Let $m \in \mathbb{N}$. The partial sum sequence of the sequence (n^m) will be denoted by $(S_m(n))$. So:

$$S_m(n) = \sum_{k=0}^{n-1} k^m.$$

11.23 Lemma. *Let (a_n) and (b_n) be sequences of numbers and let u and v be numbers. If (s_n) and (t_n) are the partial sum sequences of (a_n) and (b_n), and (d_n) and (e_n) the difference sequences, then $(us_n + vt_n)$ and $(ud_n + ve_n)$ are the partial sum sequence and the difference sequence of of the polynomial sequence $(ua_n + vb_n)$.*

Proof. This follows from

$$\sum_{k=0}^{n-1} ua_k + vb_k = u \sum_{k=0}^{n-1} a_k + v \sum_{k=0}^{n-1} b_k$$

and

$$ua_{n+1} + vb_{n+1} - ua_n - vb_n = u(a_{n+1} - a_n) + v(b_{n+1} - b_n). \qquad \square$$

11.24 Proposition. *The partial sum sequence of a polynomial sequence of degree m is a polynomial sequence of degree $m + 1$.*

Proof. We prove this by induction on m. For $m = 0$ it is clear: the partial sum sequence of the constant sequence (c) is the sequence (cn).

> Suppose the proposition holds for polynomial sequences of degree m. Then we aim to prove that it holds for polynomial sequences of degree $m + 1$ as well. By lemma 11.23 it suffices to prove this for the polynomial sequence (n^{m+1}). Let (t_n) be the sequence with $t_n = \frac{1}{m+2} n^{m+2} - S_{m+1}(n)$. The difference sequence of (t_n) is the sequence with n-th term
>
> $$\frac{1}{m+2}(n+1)^{m+2} - \frac{1}{m+2} n^{m+2} - n^{m+1}.$$
>
> This is a polynomial sequence of degree m. So (t_n) is a polynomial sequence of degree $m + 1$. From $S_{m+1}(n) = \frac{1}{m+2} n^{m+2} - t_n$ it follows that $(S_{m+1}(n))$ is a polynomial sequence of degree $m + 2$. $\square$

By lemma 11.23 the partial sum sequence of $(c_0 n^m + c_1 n^{m-1} + \cdots + c_m)$ is the sequence $(c_0 S_m(n) + c_1 S_{m-1}(n) + \cdots + c_m S_0(n))$. So the partial sum sequence is determined by the partial sum sequences $S_k(n)$ with $k \leq m$. There are $a_0, \ldots, a_{m+1}$ with $S_m(n) = a_0 n^{m+1} + a_1 n^m + \cdots + a_{m+1}$. So the $a_0, \ldots, a_{m+1}$ have to be

determined. Up to now we only know that $a_0 = \frac{1}{m+1}$ and $a_{m+1} = 0$. We will give three methods for finding the $a_0, \ldots, a_{m+1}$: Pascal's method, Newton's method and Bernoulli's method. Pascal's method gives us a recursive relation for the $S_m(n)$. Newton's method applies to any polynomial sequence, not just to (n^m). The same holds for Bernoulli's method, but applied to sequences (n^m) it will eventually lead to a recursive relation, not for the polynomials $S_m(n)$, but for their coefficients.

11.25 Notation. The partial sum sequence of a polynomial sequence $(f(n))$ is denoted by $((\Sigma f)(n))$. By theorem 11.24 it is a polynomial sequence of degree one higher than the degree of $f(n)$.

11.4.1 Pascal's method

Write the sequence (n^{m+1}) as the partial sum sequence of its difference sequence. The difference sequence of (n^{m+1}) is the sequence with n-th term

$$(n+1)^{m+1} - n^{m+1}.$$

By the binomial theorem this equals

$$\sum_{j=0}^{m} \binom{m+1}{j} n^j.$$

The sequence (n^{m+1}) is retrieved as the partial sum sequence:

$$n^{m+1} = \sum_{k=0}^{n-1} \sum_{j=0}^{m} \binom{m+1}{j} k^j = \sum_{j=0}^{m} \binom{m+1}{j} \sum_{k=0}^{n-1} k^j = \sum_{j=0}^{m} \binom{m+1}{j} S_j(n).$$

The $S_j(n)$ for $j < m$ being computed, the polynomial $S_m(n)$ can be computed from these.

11.26 Example. We compute $S_3(n)$. First we compute $S_0(n)$, $S_1(n)$ and $S_2(n)$, in this order.

$$S_0(n) = \sum_{k=0}^{n-1} k^0 = \sum_{k=0}^{n-1} 1 = n.$$

From $n^2 = \binom{2}{0} S_0(n) + \binom{2}{1} S_1(n)$ follows

$$S_1(n) = \tfrac{1}{2} n^2 - \tfrac{1}{2} S_0(n) = \tfrac{1}{2} n^2 - \tfrac{1}{2} n.$$

Next we compute $S_2(n)$:

$$n^3 = \binom{3}{0} S_0(n) + \binom{3}{1} S_1(n) + \binom{3}{2} S_2(n).$$

So

$$S_2(n) = \tfrac{1}{3}(n^3 - n - 3 \cdot \tfrac{1}{2}(n^2 - n)) = \tfrac{1}{3}n^3 - \tfrac{1}{2}n^2 + \tfrac{1}{6}n.$$

Finally

$$n^4 = \binom{4}{0} S_0(n) + \binom{4}{1} S_1(n) + \binom{4}{2} S_2(n) + \binom{4}{3} S_3(n),$$

and so

$$S_3(n) = \tfrac{1}{4}(n^4 - n - 4(\tfrac{1}{2}n^2 - \tfrac{1}{2}n) - 6(\tfrac{1}{3}n^3 - \tfrac{1}{2}n^2 + \tfrac{1}{6}n)) = \tfrac{1}{4}n^4 - \tfrac{1}{2}n^3 + \tfrac{1}{4}n^2.$$

11.4.2 Newton's method

For binomial coefficients $\binom{n}{k}$ we have

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}.$$

The right hand side is a polynomial in n. We can extend the definition of binomial coefficients.

11.27 Definition. For $x \in \mathbb{Q}$ and $k \in \mathbb{N}$ we define

$$\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}.$$

The polynomial $b_k(x) = \binom{x}{k}$ is a polynomial in x of degree k.

These more general binomial coefficients satisfy the following familiar rule:

11.28 Proposition. *For $x \in \mathbb{Q}$ and $k \in \mathbb{N}^+$ we have* $\binom{x}{k} = \binom{x-1}{k-1} + \binom{x-1}{k}$.

Proof.

$$\begin{aligned}\binom{x-1}{k-1} + \binom{x-1}{k} &= \frac{(x-1)\cdots(x-k+1)}{(k-1)!} + \frac{(x-1)\cdots(x-k)}{k!} \\ &= \frac{(x-1)\cdots(x-k+1)k}{k!} + \frac{(x-1)\cdots(x-k+1)(x-k)}{k!} \\ &= \frac{x(x-1)\cdots(x-k+1)}{k!} = \binom{x}{k}\end{aligned}$$

□

For every $k \in \mathbb{N}$ we now have a polynomial sequence $(b_k(n))$ with $b_k(n) = \binom{n}{k}$ and $b_k(n) = 0$ for $n < k$. In Figure 11.8 these zero values are added to Pascal's triangle. The rows are the polynomial sequences $(b_k(n))$.

11.29 Proposition. *For all $k \in \mathbb{N}$*

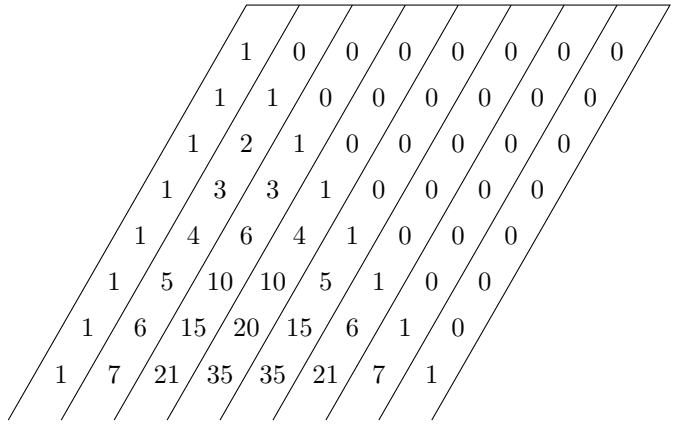


Figure 11.9: Polynomial sequences $(b_k(n))$ in Pascal's triangle

(i) $(b_k(n))$ *is the difference sequence of* $(b_{k+1}(n))$
(ii) $(b_{k+1}(n))$ *is the partial sum sequence of* $(b_k(n))$.

That is $\Delta b_{k+1} = b_k$ *and* $\Sigma b_k = b_{k+1}$.

Proof. The difference sequence of $(b_{k+1}(n))$ is $(b_{k+1}(n+1) - b_{k+1}(n))$ and by proposition 11.28 this is the sequence $(b_k(n))$. Part (ii) follows from $b_{k+1}(0) = 0$. □

11.30 Theorem (Newton). *Let* $(f(n))$ *be a polynomial sequence of degree* m. *Then for all* $n \in \mathbb{N}$:

$$f(n) = \sum_{k=0}^{m} (\Delta^k f)(0) \binom{n}{k}.$$

Proof. We prove the theorem by induction on the degree m of f. For $m = 0$ the right hand side is $f(0)$ and this is $f(n)$ since the degree is 0.

Suppose the formula is correct for polynomials of degree m. Let $f(n)$ be a polynomial of degree $m+1$. Let $g(n) = \sum_{k=0}^{m+1} (\Delta^k f)(0) \binom{n}{k}$. We prove that $(f(n)) = (g(n))$. The sequences $(f(n))$ and $(g(n))$ have the same 0-th term:

$$g(0) = \sum_{k=0}^{m+1} (\Delta^k f)(0) \binom{0}{k} = (\Delta^0 f)(0) = f(0).$$

We prove that the difference sequences are equal.

$$(\Delta g)(n) = \sum_{k=1}^{m+1} (\Delta^k f)(0) \binom{n}{k-1}$$

and since $(\Delta f)(n)$ is a polynomial of degree m, by induction hypothesis

$$(\Delta f)(n) = \sum_{k=0}^{m} (\Delta^{k+1} f)(0) \binom{n}{k}.$$

So the sequences $(f(n))$ and $(g(n))$ are equal. □

Having a polynomial sequence $f(n)$ of degree m, for a formula for $f(n)$ we only need to know $f(0), (\Delta f)(0), (\Delta^2 f)(0), \ldots, (\Delta^m f)(0)$, the formula is determined by the constant terms of the difference sequences.

11.31 Example. Again we determine the partial sum sequence of (n^3). The difference sequence of that partial sum sequence is (n^3) and the constant term of a partial sum sequence always is 0. In the following scheme there is all we need for a formula for the terms indicated by a '?'.

$$\begin{array}{ccccccccc}
0 & & ? & & ? & & ? & & ? & \cdots \\
& 0 & & 1 & & 8 & & 27 & & \\
& & 1 & & 7 & & 19 & & & \\
& & & 6 & & 12 & & & & \\
& & & & 6 & & & & &
\end{array}$$

We get

$$\begin{aligned}
S_3(n) &= 0 \cdot \binom{n}{0} + 0 \cdot \binom{n}{1} + 1 \cdot \binom{n}{2} + 6 \cdot \binom{n}{3} + 6 \cdot \binom{n}{4} \\
&= \tfrac{1}{2} n(n-1) + 6 \cdot \tfrac{1}{6} n(n-1)(n-2) + 6 \cdot \tfrac{1}{24} n(n-1)(n-2)(n-3) \\
&= \tfrac{1}{4} n^4 - \tfrac{1}{2} n^3 + \tfrac{1}{4} n^2.
\end{aligned}$$

11.4.3 Bernoulli's method

The methods in the two preceding subsections for the determination of $S_m(n)$ require a lot of computation: for Pascal's method all preceding $S_k(n)$ have to be computed first and subsequently the sum over all $\binom{m+1}{k} S_k(n)$ has to be taken, while for Newton's method the initial terms of the difference sequences have to be computed, the polynomials $\binom{n}{k}$ have to be elaborated and finally a summation has to be done.

Since a polynomial of degree $m+1$ has to be computed, one might start with an arbitrary polynomial of degree $m+1$ and compute the coefficients: the initial term is known as well as the difference sequence. That is the simplest way. For each m all this has to be done all over, unless one discovers some regularity in this process.

11.32 Example. Again the partial sum sequence of (n^3). Its degree is 4, say

$$S_3(n) = A_0n^4 + A_1n^3 + A_2n^2 + A_3n + A_4.$$

To compute $A_0, \ldots, A_4$. We already know that $A_0 = \frac{1}{4}$ and $A_4 = 0$. We compute the difference sequence (in which there is no A_4):

$$A_0(4n^3 + 6n^2 + 4n + 1) + A_1(3n^2 + 3n + 1) + A_2(2n + 1) + A_3.$$

For each n this has to be equal to n^3, that is

$$\begin{aligned} 4A_0 &= 1, \\ 6A_0 + 3A_1 &= 0, \\ 4A_0 + 3A_1 + 2A_2 &= 0, \\ A_0 + A_1 + A_2 + A_3 &= 0. \end{aligned}$$

This results from comparing coefficients. These have to be equal since otherwise there would have been an equation of degree at most 4 with more then 4 (and even infinitely many) solutions. The system of equations is easily solved from top to bottom and it is clear that there is a unique solution (what we already knew). We find subsequently $A_0 = \frac{1}{4}$, $A_1 = -\frac{1}{2}$, $A_2 = \frac{1}{4}$, $A_3 = 0$. So indeed $S_3(n) = \frac{1}{4}n^4 - \frac{1}{2}n^3 + \frac{1}{4}n^2$.

Bernoulli went even further. The description for the coefficients of $S_m(n)$ as found by Bernoulli is one that relates these coefficients for the various m. We start again as above but this time for a general m. Write

$$S_m(n) = \sum_{k=0}^{m} A_k n^{m+1-k}.$$

Since $S_m(0) = 0$, the constant term is 0. Then

$$S_m(n+1) = \sum_{k=0}^{m} A_k \sum_{l=0}^{m+1-k} \binom{m+1-k}{m+1-k-l} n^{m+1-k-l}.$$

The terms of the difference sequence are

$$n^m = S_m(n+1) - S_m(n) = \sum_{k=0}^{m} \sum_{l=1}^{m+1-k} \binom{m+1-k}{m+1-k-l} A_k n^{m+1-k-l}.$$

We collect equal powers of n and write $k + l - 1 = s$:

$$n^m = \sum_{s=0}^{m} \sum_{k=0}^{s} \binom{m+1-k}{s+1-k} A_k n^{m-s}.$$

So the A_k satisfy the $m+1$ equations

$$\sum_{k=0}^{s} \binom{m+1-k}{s+1-k} A_k = \begin{cases} 1 & \text{if } s = 0 \\ 0 & \text{if } 1 \le s \le m. \end{cases}$$

So far this is not different from what was done in the example for $m = 3$. Now for a new move. We introduce new unknowns B_k:

$$A_k = \frac{1}{m+1} \binom{m+1}{k} B_k.$$

We will see that the B_k do not depend on m. For this the following easily verifiable identity will be used

$$\binom{m+1}{k} \binom{m+1-k}{s+1-k} = \binom{m+1}{s+1} \binom{s+1}{k}.$$

For $s = 0$ we get

$$\frac{1}{m+1} \binom{m+1}{1} \binom{1}{0} B_0 = 1,$$

So $B_0 = 1$. For $s = 1, \ldots, m$:

$$\sum_{k=0}^{s} \frac{1}{m+1} \binom{m+1}{s+1} \binom{s+1}{k} B_k = 0,$$

that is

$$\sum_{k=0}^{s} \binom{s+1}{k} B_k = 0.$$

So for the B_k we have the equations

$$\begin{array}{lllllll} B_0 & & & & & = & 1, \\ B_0 & +2B_1 & & & & = & 0, \\ B_0 & +3B_1 & +3B_2 & & & = & 0, \\ B_0 & +4B_1 & +6B_2 & +4B_3 & & = & 0, \\ B_0 & +5B_1 & +10B_2 & +10B_3 & +5B_4 & = & 0, \\ & & & & & \vdots & \end{array}$$

For the formula for $S_m(n)$ only the first $m+1$ of these equations are needed.

11.33 Definition. The sequence of numbers B_n is defined by

$$\begin{cases} B_0 = 1, \\ \sum_{k=0}^{n} \binom{n+1}{k} B_k = 0 \quad \text{for all } n \in \mathbb{N}^+. \end{cases}$$

The number B_n is called the n-th *Bernoulli number*.

Jacob Bernoulli (Basel 1654 – Basel 1705)

Jacob Bernoulli came from a family of merchants. In the family there were many mathematicians. He, his younger brother **Johann** and Johann's son **Daniel** are the most important exponents. Jacob Bernoulli made influential contributions to differential calculus and he was one of the founders of stochastics, on which subject he wrote a famous book: 'Ars Conjectandi'.

From the definition it is clear that every next Bernoulli number can be calculated inside the field $\mathbb{Q}$. So the Bernoulli numbers are rational numbers.

The numbers B_n are also determined by

$$\begin{array}{rcrcrcrcrcrcl}
B_0 & & & & & & & & & & & = & B_0 \\
B_0 & + & B_1 & & & & & & & & & = & B_1 + 1 \\
B_0 & + & 2B_1 & + & B_2 & & & & & & & = & B_2 \\
B_0 & + & 3B_1 & + & 3B_2 & + & B_3 & & & & & = & B_3 \\
B_0 & + & 4B_1 & + & 6B_2 & + & 4B_3 & + & B_4 & & & = & B_4 \\
B_0 & + & 5B_1 & + & 10B_2 & + & 10B_3 & + & 5B_4 & + & B_5 & = & B_5 \\
 & & & & & & & & & & & \vdots &
\end{array}$$

Or, equivalently, we have for all $m \in \mathbb{N}$:

$$\sum_{k=0}^{m} \binom{m}{k} B_k = B_m + \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{if } m \neq 1. \end{cases}$$

The first 15 Bernoulli numbers:

$n:$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$B_n:$	1	$-\frac{1}{2}$	$\frac{1}{6}$	0	$-\frac{1}{30}$	0	$\frac{1}{42}$	0	$-\frac{1}{30}$	0	$\frac{5}{66}$	0	$-\frac{691}{2730}$	0	$\frac{7}{6}$

So we derived::

11.34 Theorem (Bernoulli). *For all $m, n \in \mathbb{N}^+$ with $n \geq 2$:*

$$S_m(n) = \frac{1}{m+1} \sum_{k=0}^{m} \binom{m+1}{k} B_k n^{m+1-k}.$$ □

11.35 Example. Again $S_3(n)$. From theorem 11.34 follows

$$\begin{aligned} S_3(n) &= \tfrac{1}{4}\left(\binom{4}{0}B_0n^4 + \binom{4}{1}B_1n^3 + \binom{4}{2}B_2n^2 + \binom{4}{3}B_3n\right) \\ &= \tfrac{1}{4}(n^4 - 4\cdot\tfrac{1}{2}n^3 + 6\cdot\tfrac{1}{6}n^2) = \tfrac{1}{4}n^4 - \tfrac{1}{2}n^3 + \tfrac{1}{4}n^2. \end{aligned}$$

The computation of the first 15 Bernoulli numbers suggests that $B_m = 0$ for m odd and ≥ 3. We will show that this is indeed the case. It is convenient to reformulate what we found in terms of Bernoulli polynomials:

11.36 Definition. Let $m \in \mathbb{N}$. The m-th *Bernoulli polynomial* $B_m(x)$ is defined as follows:

$$B_m(x) = \sum_{k=0}^{m} \binom{m}{k} B_k x^{m-k}.$$

Theorem 11.34 now becomes: $S_m(n) = \frac{1}{m+1}(B_{m+1}(n) - B_{m+1}(0))$. The polynomials $S_m(x)$ and $\frac{1}{m+1}B_{m+1}(x)$ have the same difference polynomial, namely x^m, that is the difference polynomial of $B_m(x)$ is mx^{m-1}. The constant term of $B_m(x)$ is $B_m(0) = B_m$. Moreover, from the definition of the Bernoulli numbers it follows that

$$B_m(1) = \sum_{k=0}^{m} \binom{m}{k} B_k = B_m + \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{if } m \neq 1. \end{cases}$$

We will show that $B_m = 0$ for m odd and ≥ 3. For that purpose we consider the polynomials

$$C_m(x) = (-1)^m B_m(1 - x).$$

First we compute the difference polynomial of $C_{m+1}(x)$:

$$\begin{aligned} &C_{m+1}(x+1) - C_{m+1}(x) = (-1)^{m+1}B_{m+1}(-x) - (-1)^{m+1}B_{m+1}(1-x) \\ &\quad = (-1)^m(B_{m+1}(1-x) - B_{m+1}(-x)) = (-1)^m(m+1)(-x)^m = (m+1)x^m. \end{aligned}$$

The difference polynomials of $C_{m+1}(x)$ and $B_{m+1}(x)$ are equal and therefore the polynomial $C_{m+1}(x) - B_{m+1}(x)$ is constant. In particular both polynomials have the same coefficient of x. The coefficient of x in the polynomial $B_{m+1}(x)$ is $\binom{m+1}{m}B_m = (m+1)B_m$. The coefficient of x in the polynomial $C_{m+1}(x)$ is equal to

$$\begin{aligned} (-1)^{m+1}\sum_{k=0}^{m}\binom{m+1}{k}B_k\cdot(-m-1+k) &= (-1)^m\sum_{k=0}^{m}\binom{m+1}{k}B_k\cdot(m+1-k) \\ = (-1)^m\sum_{k=0}^{m}(m+1)\binom{m}{k}B_k &= (-1)^m(m+1)\sum_{k=0}^{m}\binom{m}{k}B_k. \end{aligned}$$

So for $m \neq 1$ this coefficient is $(-1)^m(m+1)B_m$. Because it equals $(m+1)B_m$ we have shown:

Ernst Eduard Kummer (Sorau 1810 – Berlin 1893)

The German mathematician Kummer made progress in his research on Fermat's Last Theorem: he proved the theorem to be true for exponents p, where p is a 'regular' prime. A prime number p is regular if the least numerators of the Bernoulli numbers $B_k \neq 0$ for $k < p$ are no multiples of p. Unfortunately there are infinitely many irregular primes. If you do arithmetic not only with the integers, but with the integers extended with the p-th root of unity ζ_p (a complex number $\neq 1$ with $\zeta^p = 1$, see chapter 19), then in that extended system the analog of the Fundamental Theorem of Arithmetic does not hold if $p \geq 23$. However, for p regular this system nevertheless has a structure that can be used for Fermat's Last Theorem.

11.37 Proposition. *For m odd and > 1 we have $B_m = 0$.* □

From this it follows that $B_m(x)$ and $C_m(x)$ also have the same constant term: for $B_m(x)$ this is B_m and for $C_m(x)$ it is $(-1)^m B_m(1)$, which for $m \neq 1$ equals $(-1)^m B^m$. For $m = 1$ we have $B_1(x) = x - \frac{1}{2}$ and $C_1(x) = -B_1(1-x) = -(1-x-\frac{1}{2}) = x - \frac{1}{2}$. Therefore:

11.38 Proposition. *The polynomial $(-1)^m B_m(1-x)$ is the same as $B_m(x)$.* □

Python

For the computation of the n-th Bernoulli number B_n all preceding Bernoulli numbers are needed and also a new row in Pascal's triangle. The function `combrow(n)` returns the n-th row of binomial coefficients. The function `bernoulli(n)` returns the list of Bernoulli numbers up to B_n: a new row of binomial coefficients is made and subsequently used for the computation of the next Bernoulli number.

combinatorics.py

```
def combrow(n):
    m = 0
    row = [1]
    while m < n:
        row = [0] + row + [0]
        row = [row[i] + row[i + 1] for i in range(m + 2)]
        m = m + 1
    return row
```

combinatorics.py

```
from functools import reduce
from arithmetics import *
def bernoullist(n):
    if n == 0: return [(1, 1)]
    if n == 1: return [(1, 1),(-1, 2)]
    else:
        m = 1
        row = [1, 2, 1]
        bernoul = [(1, 1), (-1, 2)]
        while m < n:
            m = m + 1
            row = [0] + row + [0]
            row = [row[i] + row[i + 1] for i in range(m + 2)]
            bernnew = [mul((row[i], 1), bernoul[i]) for i in\
            range(m)]
            bernoul.append(mul(reduce(add, bernnew),(-1, m + 1)))
        return bernoul

def bernoulli(n):
    return bernoullist(n)[n]
```

```
>>> combrow(32)
[1, 32, 496, 4960, 35960, 201376, 906192, 3365856, 10518300, 28048800
, 64512240, 129024480, 225792840, 347373600, 471435600, 565722720, 60
1080390, 565722720, 471435600, 347373600, 225792840, 129024480, 64512
240, 28048800, 10518300, 3365856, 906192, 201376, 35960, 4960, 496, 3
2, 1]
>>> bernoullist(20)
[(1, 1), (-1, 2), (1, 6), (0, 1), (-1, 30), (0, 1), (1, 42), (0, 1),
(-1, 30), (0, 1), (5, 66), (0, 1), (-691, 2730), (0, 1), (7, 6), (0,
1), (-3617, 510), (0, 1), (43867, 798), (0, 1), (-174611, 330)]
>>> [i[0]%37 for i in bernoullist(37)]
[1, 36, 1, 0, 36, 0, 1, 0, 36, 0, 5, 0, 12, 0, 7, 0, 9, 0, 22, 0, 29
, 0, 35, 0, 12, 0, 35, 0, 33, 0, 6, 0, 0, 0, 12, 0, 4, 0]
>>> bernoulli(32)
(-7709321041217, 510)
>>> divmod(-7709321041217, 37)
(-208360028141, 0)
>>> divmod(bernoulli(44)[0], 59)
(-471750331852559732797, 0)
```

`[i[0]%37 for i in bernoulli(37)]` returns the remainder after division by 37 of the numerators of the B_k with $k \leq 37$. The prime number 37 is the least irregular prime. The next one is 59.

11.5 The Inclusion-Exclusion Principle

The Inclusion-Exclusion Principle is applicable to many counting problems. In section 12.1 we will apply it to the counting of derangements, permutations with no fixed point. In the next section the principle will be used for deriving a formula for the number of surjections between finite sets, a number related to the number of partitions of a finite set with a given number of classes.

Let be given:

- a finite set A,
- a finite index set I,
- for each $i \in I$ a subset A_i of A.

The Inclusion-Exclusion Principle is used for determination of the number of elements of A not lying in one of the subsets A_i. The formula for this number contains numbers of elements of intersections of subsets A_i only. For convenience we introduce a short notation for these intersections. For $J \subseteq I$ we write

$$A_J = \begin{cases} \bigcap_{j \in J} A_j & \text{if } J \neq \emptyset, \\ A & \text{if } J = \emptyset. \end{cases}$$

The $\bigcap$-notation has the following meaning:

$$\bigcap_{i \in I} A_i = \{\, a \in A \mid a \in A_i \text{ for all } i \in I \,\}.$$

So $A_\emptyset = A$, $A_{\{i\}} = A_i$, $A_{\{i,j\}} = A_i \cap A_j$, $A_{\{i,j,k\}} = A_i \cap A_j \cap A_k$, etc. This notation is similar to the $\sum$-notation. Here too we have an abelian monoid: the set $\mathcal{P}(A)$ together with the operation $\cap$; the neutral element is A itself. For the $\bigcap$-notation the index set I is not necessarily finite: any collection of sets has an intersection. Analogous to the $\bigcap$-notation there is a $\bigcup$-notation for the union of a collection of sets. This section is about counting and therefore, sets are finite.

We will use characteristic functions of subsets of A, see the sections 5.9 and 8.2. The behavior of characteristic functions under taking complements and intersections is simple:

11.39 Lemma. *Let U and V be subsets of A. Then for all $a \in A$:*

$$\chi_{A \setminus U}(a) = 1 - \chi_U(a) \quad \textit{and} \quad \chi_{U \cap V}(a) = \chi_U(a)\chi_V(a). \qquad \square$$

For the characteristic functions we will use short notations:

$$\chi_i = \chi_{A_i} \quad \text{and} \quad \chi_J = \chi_{A_J}.$$

The Inclusion-Exclusion Principle yields a formula for $\#(A \setminus \bigcup_{i\in I} A_i)$. From $A \setminus \bigcup_{i\in I} A_i = \bigcap_{i\in I}(A \setminus A_i)$ follows

$$\chi_{A\setminus\bigcup_{i\in I} A_i}(a) = \chi_{\bigcap_{i\in I}(A\setminus A_i)}(a) = \prod_{i\in I} \chi_{A\setminus A_i}(a) = \prod_{i\in I}(1-\chi_i(a)).$$

We apply theorem 11.12:

$$\chi_{A\setminus\bigcup_{i\in I} A_i}(a) = \sum_{J\subseteq I}(-1)^{\#(J)}\chi_J(a).$$

Summation over $a \in A$ yields:

$$\#\Big(A \setminus \bigcup_{i\in I} A_i\Big) = \sum_{J\subseteq I}(-1)^{\#(J)} \sum_{a\in A} \chi_J(a) = \sum_{J\subseteq I}(-1)^{\#(J)}\#(A_J).$$

We introduce some standard notations:

$$\begin{aligned} N &= \#(A), \\ N_i &= \#(A_i), \\ N_I &= \#(A_I), \\ S_k &= \sum_{\#(I)=k} N_I. \end{aligned}$$

Instead of for example $N_{\{1,2\}}$ we will usually write $N_{1,2}$. We have: $S_0 = N_\emptyset = \#(A_\emptyset) = \#(A) = N$. The Inclusion-Exclusion Principle can now be formulated as follows:

11.40 Theorem (Inclusion-Exclusion Principle). *Using the notations in this section: the number of elements of A not in one of the subsets A_i equals*

$$\sum_{k=0}^{n}(-1)^k S_k. \qquad \square$$

11.41 Example. We will determine how many of the numbers in $\underline{100}$ are not divisible by 2, 3, 5 or 7. We use the index set $\{2,3,5,7\}$. Let A_2, A_3, A_5 and A_7 be the subsets of $A = \underline{100}$ of the multiples of respectively 2, of 3, of 5 and of 7. We have $S_0 = N = 100$ and

$$\begin{aligned} N_2 &= \lfloor \tfrac{100}{2} \rfloor = 50 \\ N_3 &= \lfloor \tfrac{100}{3} \rfloor = 33 \\ N_5 &= \lfloor \tfrac{100}{5} \rfloor = 20 \\ N_7 &= \lfloor \tfrac{100}{7} \rfloor = 14, \end{aligned}$$

and so $S_1 = 50 + 33 + 20 + 14 = 117$. Since the four numbers 2, 3, 5 and 7 are pairwise relatively prime, we have

$$\begin{aligned} N_{2,3} &= \lfloor \tfrac{100}{6} \rfloor = 16 \\ N_{2,5} &= \lfloor \tfrac{100}{10} \rfloor = 10 \\ N_{2,7} &= \lfloor \tfrac{100}{14} \rfloor = 7 \\ N_{3,5} &= \lfloor \tfrac{100}{15} \rfloor = 6 \\ N_{3,7} &= \lfloor \tfrac{100}{21} \rfloor = 4 \\ N_{5,7} &= \lfloor \tfrac{100}{35} \rfloor = 2, \end{aligned}$$

and so $S_2 = 16 + 10 + 7 + 6 + 4 + 2 = 45$.

$$\begin{aligned} N_{2,3,5} &= \lfloor \tfrac{100}{30} \rfloor = 3 \\ N_{2,3,7} &= \lfloor \tfrac{100}{42} \rfloor = 2 \\ N_{2,5,7} &= \lfloor \tfrac{100}{70} \rfloor = 1 \\ N_{3,5,7} &= \lfloor \tfrac{100}{105} \rfloor = 0. \end{aligned}$$

So $S_3 = 3+2+1 = 6$. From $2 \cdot 3 \cdot 5 \cdot 7 = 210 > 100$ it follows that $S_4 = 0$. So there there are $100 - 117 + 45 - 6 = 22$ numbers in $\underline{100}$ which are not a multiple of 2, 3, 5 or 7. Thus this number has been calculated without looking at the individual numbers separately.

11.6 Surjective Maps and Partitions

11.6.1 Surjective maps

We will derive a formula for the number of surjective maps from $\underline{n}$ to $\underline{k}$.

Let A be the set of all maps from $\underline{n}$ to $\underline{k}$. We will use the Inclusion-Exclusion Principle. As index set we take $\underline{k}$. Let for $i \in \underline{k}$ the subset A_i of A consist of all $f\colon \underline{n} \to \underline{k}$ with $i \notin f_*(\underline{n})$. For $J \subseteq \underline{k}$ we have $N_J = (k - \#(J))^n$ and, because there are $\binom{k}{j}$ subsets of $\underline{k}$ with j elements, we have $S_j = \binom{k}{j}(k-j)^n$. So we have:

11.42 Proposition. *The number of surjective maps from a set of n elements to a set of k elements is*

$$\sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n. \qquad \square$$

11.43 Example. At a bingo party 20 people participate and there are 25 prizes. What is the probability that everybody wins a prize? This chance is given by the

ratio of the number of surjective maps from $\underline{25}$ to $\underline{20}$ and the total number of maps. The number of surjections is:

$$\binom{20}{0}20^{25} - \binom{20}{1}19^{25} + \cdots + (-1)^j\binom{20}{j}(20-j)^{25} + \cdots + \binom{20}{20}0^{25}$$

and the total number of maps is 20^{25}. A computer is needed for the calculation of this number. Also have a look at example 11.51.

11.6.2 Partitions with k classes

11.44 Definition. Let n, k be natural numbers. We define the *Stirling number* $\left\{{n \atop k}\right\}$ of the *second kind* as the number of partitions Φ of a set of n elements satisfying $\#(\Phi) = k$.

11.45 Example. The partitions of $\underline{4}$ with two classes are:

$$\begin{array}{lll} \{\{1,2\},\{3,4\}\}, & \{\{1,3\},\{2,4\}\}, & \{\{1,4\},\{2,3\}\}, \\ \{\{1\},\{2,3,4\}\}, & \{\{2\},\{1,3,4\}\}, & \{\{3\},\{1,2,4\}\}, \\ \{\{4\},\{1,2,3\}\} & & \end{array}$$

So $\left\{{4 \atop 2}\right\} = 7$.

> The Scottish mathematician **James Stirling** (Garden 1692 – Edinburgh 1770) was a contemporary of Euler. In Newton's tradition he was working in the area of calculus. He is known because of Stirling's formula
>
> $$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$
>
> an approximation of $n!$ probably obtained earlier by the French mathematician **Abraham de Moivre** (Vitry-le-François 1667 – Londen 1754) who, after he fled France, was mainly active in England.

The Stirling numbers can be displayed in a triangle in the same way as the binomial coefficients. The following recursive description can be used for their computation.

11.46 Theorem. *Stirling numbers of the second kind satisfy:*

(i) $\left\{{n \atop 0}\right\} = 0$ *for all* $n \in \mathbb{N}^+$,

(ii) $\left\{{n \atop n}\right\} = 1$ *for all* $n \in \mathbb{N}$,

(iii) $\left\{{n \atop k}\right\} = \left\{{n-1 \atop k-1}\right\} + k\left\{{n-1 \atop k}\right\}$ *for all* $n, k \in \mathbb{N}^+$ *with* $0 < k < n$.

Proof. The first two parts are simple. For the last part choose a fixed element a in a set A of n elements. The partitions with k classes come in two kinds:

				1				
			0		1			
		0		1		1		
	0		1		3		1	
0		1		7		6		1
	0		1		15		25		10		1
		0		1		31		90		65		15		1
	0		1		63		301		350		140		21		1
0		1		127		966		1701		1050		266		28		1

Figure 11.10: Triangle of Stirling numbers of the second kind

a) partitions Φ with $\{a\} \in \Phi$; of these there are $\left\{ {n-1 \atop k-1} \right\}$,
b) partitions Φ with $\{a\} \notin \Phi$; these correspond to partitions of $A \setminus \{a\}$ with k classes together with a choice of one of these classes: the class to which a is to be added; of these there are $k\left\{ {n-1 \atop k} \right\}$. □

For small n we get the numbers as displayed in Figure 11.10.

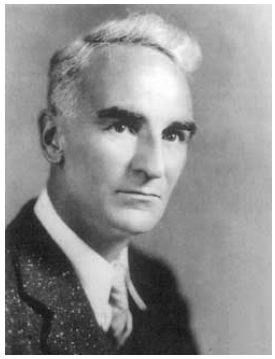
Python

The Python code is analogous to the code for the computation of binomial coefficients.

combinatorics.py

```
def stirling2(n, k):
    list0 = [1] + (n - k) * [0]
    i = j = 0
    while i < k:
        list1 = [1]
        i = i + 1
        while j < n - k:
            j = j + 1
            list1.append(i * list1[-1] + list0[j])
        list0 = list1
        j = 0
    return list0[-1]
```

Eric Temple Bell (Peterhead, Scotland 1883 – Watsonville, USA 1960)



The Bell numbers are named after Eric Temple Bell, who is known as a writer of books on the history of mathematics. Under the name John Taine he also wrote science fiction.

```
>>> stirling2(153, 60)
616578310862199096208547253227976406265353676270629543580059477056658
923549136823575833236201554312588011711962563493018153940980871150296
155555724403504410288600020770540292195664245884 00
```

In the triangle of Stirling numbers of the second kind row n consists of all numbers of partitions of $\underline{n}$ with a given number of classes.

11.47 Definition. The n-th *Bell number* b_n is the number of partitions of a set which has n elements.

So by definition:

11.48 Proposition. *For every $n \in \mathbb{N}$ we have $b_n = \sum_{k=0}^{n} \left\{ {n \atop k} \right\}$.* □

11.49 Example. The Bell number can be computed from the Stirling numbers of the second kind, see Figure 11.10:

$$
\begin{aligned}
b_0 &= 1 \\
b_1 &= 0 + 1 = 1 \\
b_2 &= 0 + 1 + 1 = 2 \\
b_3 &= 0 + 1 + 3 + 1 = 5 \\
b_4 &= 0 + 1 + 7 + 6 + 1 = 15 \\
b_5 &= 0 + 1 + 15 + 25 + 10 + 1 = 52 \\
b_6 &= 0 + 1 + 31 + 90 + 65 + 15 + 1 = 203 \\
b_7 &= 0 + 1 + 63 + 301 + 350 + 140 + 21 + 1 = 877
\end{aligned}
$$

A surjective map from $\underline{n}$ to $\underline{k}$ determines a partition of $\underline{n}$ into k classes. Such a partition Φ comes from $k!$ of such surjective maps: these correspond to bijections $\underline{n}/\Phi \to \underline{k}$. So from proposition 11.42 follows:

11.50 Proposition. *For $k, n \in \mathbb{N}$*

$$k! \cdot \left\{ {n \atop k} \right\} = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k-j)^n. \qquad \square$$

11.51 Example. We return to example 11.43. The number of surjections from $\underline{25}$ to $\underline{20}$ is $20! \cdot \left\{ {25 \atop 20} \right\}$ and this last expression is more easily evaluated than the sum found in 11.43.

```
>>> factorial(20) * stirling2(25,20)
151331242985247932006400000000
```

The probability asked for in example 11.43 is approximately 0.000045:

```
>>> (factorial(20) * stirling2(25, 20)) / pow(20, 25)
4.5100224907770134e-05
```

Exercises

1. Prove proposition 11.7 and proposition 11.8 using the formula $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
2. Let $n \in \mathbb{N}^+$. Prove that $n^n \geq n!\, 2^{n-1}$.
3. Prove that for all $n \in \mathbb{N}$
$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$
4. Determine a formula for the sum of the 5th powers of the first n natural numbers. Do so using Pascal's method, Newton's method and Bernoulli's method.
5. Let p be a prime number and $k \in \mathbb{N}$ with $0 < k < p$. Prove that $p \mid \binom{p}{k}$.
6. Prove Fermat's Little Theorem: $p \mid n^p - n$ for all prime numbers p and all $n \in \mathbb{N}$. Prove this theorem by mathematical induction on n. Use exercise 5. (There is another proof in chapter 13: it is Corollary 13.23.)
7. Prove that for all $m, n \in \mathbb{N}$
$$\sum_{k=0}^{n} \binom{m+k}{k} = \binom{m+n+1}{n}.$$

8. Prove that for all $n \in \mathbb{N}$

$$f_{n+1} = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n-k}{k}.$$

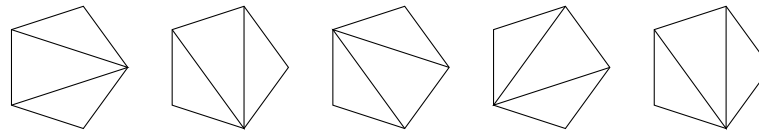
(The f_n are the Fibonacci numbers.)

9. Prove that for all $n \in \mathbb{N}$

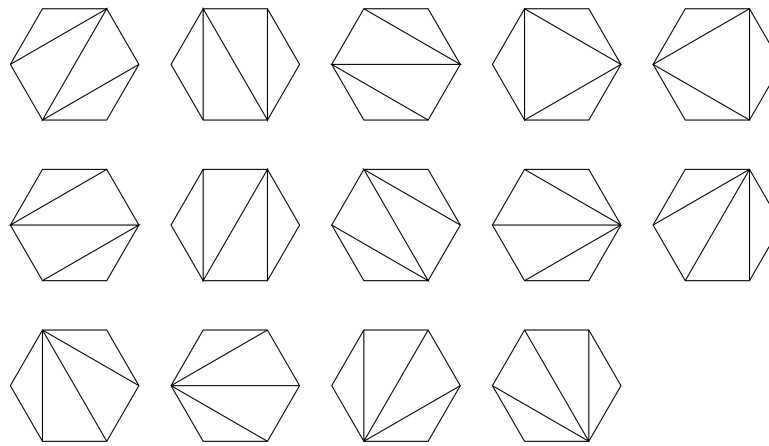
$$c_n = \frac{2 \cdot 6 \cdot 10 \cdots (4n-2)}{(n+1)!}.$$

(c_n is the nth Catalan number.)

10. Find a formula for the number of ways an n-gon (without reflex angles) can be divided in triangles by nonintersecting diagonals. Prove the formula. For a 5-gon there are 5 ways:



And for a 6-gon the number is 14:



11. For a nonassociative operation $(a, b) \mapsto ab$ in a set A the expressions $(ab)c$ and $a(bc)$ can have different meanings. In a word of 3 letters, parentheses can be placed in 2 ways. In a word of length 4 it can be done in 5 ways:

$$a(b(cd)), \quad a((bc)d), \quad (ab)(cd), \quad ((ab)c)d \quad \text{and} \quad (a(bc))d.$$

The problem is: in how many ways can it be done in a word of length n? Give a proof.

12. How many of the 26! orders of the 26 letters of the alphabet do not contain the following combinations: `kim`, `john` and `nigel`?

13. How many of the natural numbers n with $1000 \leq n \leq 10000$ are no multiple of 2, 3, 5 or 7?

14. Let $n \in \mathbb{N}^+$.
    (i) Let $d \mid n$. Show that the number of $a \in \mathbb{N}_n$ satisfying $a \mid n$ is equal to $\frac{n}{d}$.
    (ii) Let P be the set of the prime divisors of n. Show that
    $$\gcd(a, n) = 1 \iff p \nmid a \text{ for all } p \in P.$$
    (iii) Use the Inclusion-Exclusion Principle to determine a formula for $\varphi(n)$, the totient of n. Show that this formula agrees with $\varphi = \mu * \mathrm{id}$.
    (iv) Derive the formula $\varphi(n) = n \prod_{p|n}(1 - \frac{1}{p})$ from the previous part. Use theorem 11.12.

15. Show that $\sum_{j=0}^{n}(-1)^j \binom{n}{j}(n-j)^n = n!$.

16. We define polynomials $x^{\underline{n}}$ as follows
    $$x^{\underline{n}} = x(x-1)(x-2)\dots(x-n+1).$$
    Or more precisely:
    $$\begin{cases} x^{\underline{0}} = 1, \\ x^{\underline{n+1}} = x^{\underline{n}} \cdot (x-n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$
    Prove that for all $n \in \mathbb{N}$
    $$x^n = \sum_{k=0}^{n} \left\{ n \atop k \right\} x^{\underline{k}}.$$

17. Find a polynomial $f(x)$ satisfying $f(x+1) - f(x) = x^3 - 3x^2 + 1$.

18. (i) Show that $\left\{ n \atop 1 \right\} = 1$ for all $n \in \mathbb{N}$.
    (ii) Show that $\left\{ n+1 \atop 2 \right\} = 2^n - 1$ for all $n \in \mathbb{N}^+$.
    (iii) Show that $\left\{ n \atop n-1 \right\}$ is a triangular number for all $n \in \mathbb{N}^+$.

19. Let $k \in \mathbb{N}$. Is the sequence $(\left\{ n+k \atop n \right\})$ a polynomial sequence?

20. For $m \in \mathbb{N}$ the polynomial $D_m(x)$ is defined as follows
    $$D_m(x) = 2^{m-1}(B_m(\tfrac{x}{2}) + B_m(\tfrac{x+1}{2})).$$
    (i) Show that $D_m(x)$ and $B_m(x)$ have the same difference polynomial.
    (ii) Show that $D_m(x) = B_m(x)$.
    (iii) Show that $B_m(\frac{1}{2}) = -(1 - 2^{1-m})B_m$.

21. Let $n \in \mathbb{N}^+$. The following numbers are given

$$A = \sum_{k=0}^{n} \binom{2n}{2k} \quad \text{and} \quad B = \sum_{k=0}^{n-1} \binom{2n}{2k+1}.$$

Determine A and B. (Hint: look at $A + B$ and $A - B$.)

22. Let $n \in \mathbb{N}^+$. Prove that

$$\sum_{I \in \mathcal{P}(\underline{n})} \#(\mathcal{P}(I)) = 3^n.$$

12 Permutations

In this chapter we consider permutations of finite sets. Some parts of this chapter will have applications in chapter 13, where we will work in finite rings. Multiplication by an invertible element in a finite ring is an example of such a permutation. Permutations occur in many parts of mathematics, in particular in group theory. The section on the sign of a permutation is also of interest for the theory of determinants in linear algebra.

12.1 Orbits

Let σ be a permutation of a set A. In particular σ is a transformation and, therefore, we also have iterates of σ, the transformations σ^n with $n \in \mathbb{N}$. Because σ is bijective, we also have the permutation σ^{-1}, the inverse of σ. More generally we have:

12.1 Definition. Let σ be a permutation of a set A. Permutations σ^n are defined for all $n \in \mathbb{Z}$ by

$$\sigma^n = \begin{cases} \sigma^n & \text{for } n \geq 0 \\ (\sigma^{-1})^{-n} & \text{for } n < 0. \end{cases}$$

As for exponentiation in $\mathbb{Q}^*$, the usual rules are satisfied. The proofs are not different from those we presented in subsection 9.1.4 for this exponentiation. In fact, all this is applicable to any group; the permutations of a set form a group, the operation being the composition of permutations.

12.2 Proposition. *Let σ and τ be permutations of a set A satisfying $\sigma\tau = \tau\sigma$ and let m and n be integers. Then:*

(i) $\sigma^m\sigma^n = \sigma^{m+n}$,
(ii) $(\sigma^m)^n = \sigma^{mn}$,
(iii) $(\sigma\tau)^n = \sigma^n\tau^n$. □

12.3 Definition. Let σ be a permutation of a set A. The set

$$[a]_\sigma = \{\, \sigma^n(a) \mid n \in \mathbb{Z} \,\}$$

is called the *orbit* of a under σ. We also say that it is an *orbit* of σ. The set of all orbits of σ we denote by A_σ.

12.4 Example. Figure 12.1 is a picture of

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 5 & 4 & 3 & 6 & 9 & 7 & 8 & 1 \end{pmatrix}.$$

There are 4 orbits:
$\{1, 2, 5, 6, 9\}$, $\{3, 4\}$, $\{7\}$ and $\{8\}$.

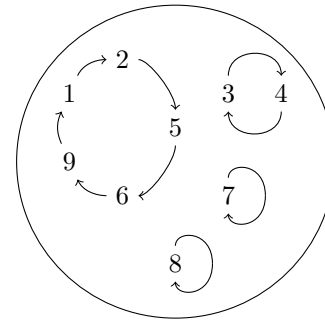


Figure 12.1: Picture of a permutation

A permutation σ of A determines an equivalence relation in A, the equivalence classes are the orbits of σ.

12.5 Definition. Let σ be a permutation of A. Then a relation $\sim_\sigma$ is defined by

$$a \sim_\sigma b \iff \text{there is an } n \in \mathbb{Z} \text{ with } \sigma^n(a) = b.$$

12.6 Proposition. *The relation $\sim_\sigma$ is an equivalence relation. The equivalence classes are the orbits of σ.*

Proof.

Reflexivity: $\sigma^0(a) = a$.

Symmetry: if $\sigma^n(a) = b$, then $\sigma^{-n}(b) = a$.

Transitivity: if $\sigma^n(a) = b$ and $\sigma^m(b) = c$, then $\sigma^{m+n}(a) = c$.

The equivalence class containing the element a is

$$\{\, x \in A \mid x \sim_\sigma a \,\} = \{\, \sigma^n(a) \mid n \in \mathbb{Z} \,\} = [a]_\sigma. \qquad \square$$

12.2 Cycles

Cycles are permutations of a special kind:

12.7 Definition. Let A be a set and let $(a_1, a_2, \ldots, a_n)$ be a finite sequence of different elements of A. We define a permutation σ of A by:

$$\begin{cases} \sigma(a_k) = a_{k+1} & \text{for } 1 \le k < n \\ \sigma(a_n) = a_1 & \\ \sigma(x) = x & \text{for all } x \in A \setminus \{a_1, a_2, \ldots, a_n\}. \end{cases}$$

Such a permutation is called an *n-cycle*. Notation: $\sigma = (a_1 \quad a_2 \quad \cdots \quad a_n)$.

The notation does not show which set is involved: the elements of A which do not occur in the notation do not occur and only from the context can it be clear which are these elements. A 1-cycle is the identical permutation: $(a) = 1_A$ for all $a \in A$. So there is a unique 1-cycle, it is the identity. Note that $(a) = (b)$ for any $a, b \in A$. Also note that $(a_1 \quad a_2 \quad \cdots \quad a_n) = (a_2 \quad \cdots \quad a_n \quad a_1)$.

Let σ be a permutation of a set A. An orbit $[a]_\sigma$ of m elements determines an m-cycle τ of A:

$$\tau(x) = \begin{cases} \sigma(x) & \text{if } x \in [a]_\sigma, \\ x & \text{otherwise.} \end{cases}$$

It is the cycle $(a \quad \sigma(a) \quad \sigma^2(a) \quad \cdots \quad \sigma^{m-1}(a))$. Restricted to the orbit of a it coincides with σ, the elements outside the orbit are mapped to themselves.

12.8 Definition. Let σ be a permutation of a set A. The set

$$D(\sigma) = \{\, a \in A \mid \sigma(a) \neq a \,\}$$

is called the *support* of the permutation σ.

The support of a n-cycle with $n > 1$ has n elements; it is the unique orbit with more than one element. In general the support is the union of the orbits with more than one element. A 1-cycle is the identical permutation; it has an empty support.

12.9 Definition. Permutations σ and τ of a set A are said to be *disjoint* if their supports are disjoint: $D(\sigma) \cap D(\tau) = \emptyset$.

12.10 Proposition. *Let σ and τ be disjoint permutations of a set A. Then $\sigma\tau = \tau\sigma$.*

Proof. For every $a \in A$ either $a \notin D(\sigma)$ or $a \notin D(\tau)$. In both cases it is easily shown that $\sigma(\tau(a)) = \tau(\sigma(a))$. □

If σ is a permutation of a finite set A with k orbits, then σ is a product of k disjoint cycles. Orbits of one element determine 1-cycles and these may be left out in the product.

12.11 Example. The permutation σ of example 12.4 has four orbits, two of which have more than one element. It is a product of a 5-cycle and a 2-cycle:

$$\sigma = (1 \quad 2 \quad 5 \quad 6 \quad 9)(3 \quad 4)(7)(8) = (1 \quad 2 \quad 5 \quad 6 \quad 9)(3 \quad 4).$$

Python

We consider only permutations of the standard sets $\underline{n}$. A permutation of $\underline{n}$ is represented by a list of the numbers 1 up to n. The image of i is the number with index $i-1$. In Python numbering starts with 0. The factorization of a permutation as a product of disjoint cycles is easy. The result is a list of disjoint cycles. Cycles are here represented by tuples. We keep the 1-cycles in the notation: they appear as `(a,)`.

combinatorics.py

```
def cycledecomposition(perm):
    sublist = perm[:]
    cycles = []
    while sublist != []:
        i = sublist[0]
        cycle=(i, )
        j = perm[i-1]
        del sublist[0]
        while j != i:
            cycle = cycle + (j, )
            sublist.remove(j)
            j = perm[j - 1]
        cycles.append(cycle)
    return cycles
```

```
>>> cycledecomposition([12, 1, 13, 5, 4, 9, 7, 6, 3, 2, 8, 10, 11])
[(12, 10, 2, 1), (13, 11, 8, 6, 9, 3), (5, 4), (7,)]
```

Codes for the conversion from the cycle notation to the standard notation and for the composition of permutations:

combinatorics.py

```
def permutation1(cycle, n):
    i = 0
    perm = []
    cyclelist = list(cycle)
    while i < n:
        i = i + 1
        if i in cycle and i != cycle[-1]:
            perm.append(cycle[cyclelist.index(i) + 1])
        elif i == cycle[-1]:
            perm.append(cycle[0])
        else:
            perm.append(i)
    return perm

def composition(perm1, perm2):
    return [perm1[perm2[i] - 1] for i in range(len(perm2))]
```

combinatorics.py
```
def permutation(cycles, n):
    return reduce(composition, [permutation1(cycle, n) for
    cycle in cycles])
```

```
>>> permutation1((4, 5, 6, 2, 7, 1), 9)
[4, 7, 3, 5, 6, 2, 1, 8, 9]
>>> composition([1, 5, 9, 4, 2, 8, 7, 2, 3], [8, 4, 3, 2, 9, 7, 6, 5,
 1])
[2, 4, 9, 5, 3, 7, 8, 2, 1]
>>> permutation([(4, 5, 6, 7, 1), (2, 3), (8, 9)], 9)
[4, 3, 2, 5, 6, 7, 1, 9, 8]
>>> permutation([(4, 5, 6, 7, 1), (2,3 ), (8, 9), (1, 7, 8, 4, 2)], 9
)
[1, 4, 2, 3, 6, 7, 9, 5, 8]
```

12.3 Derangements

For a Dutch Sinterklaas-party n persons, numbered 1 up to n, draw lots with (the names of) 1 up to n. This determines a permutation σ of $\underline{n}$: $\sigma(i) = j$ if person j draws lot i. The idea is that at the party person j has a gift for person $\sigma(j)$. Obviously, the procedure succeeds when $\sigma(j) \neq j$ for all j. The permutation then has no one element orbit. What is the chance for the procedure to succeed?

12.12 Definition. A permutation without fixed points is called a *derangement*.

We will use the inclusion-exclusion principle for finding the number of derangements. The set A is the set of all permutations of $\underline{n}$. As index set we take the set $\underline{n}$. For every $j \in \underline{n}$ we have the subset A_j of all σ with $\sigma(j) = j$.

The total number of permutations is $\#(A) = n!$. The number of derangements is $\#(A \setminus \bigcup_{j=1}^{n} A_j)$.

For $J \subseteq \underline{n}$ the elements of A_J correspond to permutations of $\underline{n} \setminus J$ and so $N_J = (n - \#(J))!$. There are $\binom{n}{k}$ subsets J of $\underline{n}$ with $\#(J) = k$ and so $S_k = \binom{n}{k}(n-k)! = \frac{n!}{k!}$. So we derived:

12.13 Proposition. *The number of derangements of a set of n elements is*

$$n!\Big(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!}\Big). \qquad \square$$

So the chance that the procedure of drawing lots for the Sinterklaas-party is successful equals

$$1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!}.$$

So the chance tends to $\frac{1}{e}$, where e is the base of the natural logarithm, see section 17.6. For even n it is larger, for odd n it is less.

12.4 Permutations with k Orbits

In the previous chapter Stirling numbers of the second kind were introduced. Now a definition of Stirling numbers of the first kind:

12.14 Definition. Let n, k be natural numbers. We define the *Stirling number* $\begin{bmatrix} n \\ k \end{bmatrix}$ of the *first* kind as the number of permutations σ of a set A of n elements which have k orbits, that is $\#(A_\sigma) = k$.

12.15 Example. The permutations of $\underline{4}$ with 2 orbits are:

$(1\ \ 2)(3\ \ 4)$	$(1\ \ 3)(2\ \ 4)$	$(1\ \ 4)(2\ \ 3)$	$(2\ \ 3\ \ 4)$
$(2\ \ 4\ \ 3)$	$(1\ \ 3\ \ 4)$	$(1\ \ 4\ \ 3)$	$(1\ \ 2\ \ 4)$
$(1\ \ 4\ \ 2)$	$(1\ \ 2\ \ 3)$	$(1\ \ 3\ \ 2)$	

So $\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 11$.

From the following theorem it follows that also these Stirling numbers can be computed recursively.

12.16 Theorem. *Stirling numbers of the first kind satisfy:*

(i) $\begin{bmatrix} n \\ 0 \end{bmatrix} = 0$ *for all* $n \in \mathbb{N}^+$,

(ii) $\begin{bmatrix} n \\ n \end{bmatrix} = 1$ *for all* $n \in \mathbb{N}$,

(iii) $\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} + (n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix}$ *for all* $n, k \in \mathbb{N}^+$ *such that* $0 < k < n$.

Proof. The first two parts are obvious. For the last part choose an element a in a set A of n elements. There are two kinds of permutations of A with k orbits:

a) permutations σ satisfying $\{a\} \in A_\sigma$; of these there are $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}$;
b) permutations σ satisfying $\{a\} \notin A_\sigma$; these correspond to permutations of $A \setminus \{a\}$ with k orbits together with the placement of the element a in one of these k orbits (there are $n-1$ ways for this); in total there are $(n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix}$ of these permutations. □

Using this theorem the Stirling numbers in Figure 12.2 are easily computed.

1

0 1

0 1 1

0 2 3 1

0 6 11 6 1

0 24 50 35 10 1

0 120 274 225 85 15 1

0 720 1764 1624 735 175 21 1

0 5040 13068 13132 6769 1960 322 28 1

Figure 12.2: Triangle of Stirling numbers of the first kind

Python

The computation of these Stirling numbers is done in the same way as the computation of the Stirling numbers of the second kind. The code for the Stirling numbers of the second kind is easily adapted for this purpose. We add it to the module `combinatorics.py`.

combinatorics.py

```
def stirling1(n, k):
    list0 = [1] + (n - k) * [0]
    i = j = 0
    while i < k:
        list1 = [1]
        i = i + 1
        while j < n - k:
            j = j + 1
            list1.append((i + j - 1) * list1[-1] + list0[j])
        list0 = list1
        j = 0
    return list0[-1]
```

```
>>> stirling1(200, 75)
33268749272775607751407633172770287842357372092135253132048795482419782168689049607621733471420840301754723156673420422196493649596357758207058667593297874727547260345133493497156555719911394693686320984913481981072357429385855454861485574220209087794254178322775573351811274553312724268770918400000
```

12.5 The Sign of a Permutation

Let A be a finite set and let a and b be elements of A such that $a \neq b$. The *transposition* $\tau_{a,b}$, see definition 5.19, of a and b is the 2-cycle $(a \quad b)$. It is a permutation of A.

12.17 Definition. Let σ be a permutation of a finite set A. The *sign* of σ, $\operatorname{sgn}(\sigma)$, is defined as

$$\operatorname{sgn}(\sigma) = (-1)^{\#(A)-\#(A_\sigma)},$$

where A_σ is the set of orbits of σ, see definition 12.3. The permutation σ is called *even* if $\operatorname{sgn}(\sigma) = 1$, so if $\#(A) - \#(A_\sigma)$ is even. If $\operatorname{sgn}(\sigma) = -1$, that is $\#(A) - \#(A_\sigma)$ is odd, then σ is called an *odd* permutation.

12.18 Lemma. *Transpositions are odd permutations.*

Proof. The number of orbits of a transposition is one less than the number of elements of the set. □

12.19 Example. The permutation $\sigma = (1 \quad 2 \quad 4)(3 \quad 5)$ of $\underline{5}$ is a permutation of a set of 5 elements and it has 2 orbits, so $\operatorname{sgn}(\sigma) = (-1)^{5-2} = -1$. It is an odd permutation.

12.20 Lemma. *Let σ be a permutation of a finite set A, and let a and b be elements of A. Then*

$$\#(A_{(a\ b)\sigma}) = \begin{cases} \#(A_\sigma) + 1 & \text{if } a \sim_\sigma b \\ \#(A_\sigma) - 1 & \text{if not } a \sim_\sigma b. \end{cases}$$

Proof.

Suppose a and b are in the same orbit of σ, say this orbit consists of the following n elements of A:

$$a, \sigma(a), \sigma^2(a), \ldots, \sigma^k(a)(= b), \ldots, \sigma^{n-1}(a)$$

(with $0 < k < n$). Then the orbit of a under $(a \quad b)\sigma$ is

$$\{a, \sigma(a), \ldots, \sigma^{k-1}(a)\},$$

and the orbit of b under $(a \quad b)\sigma$ is

$$\{\sigma^k(a)(= b), \sigma^{k+1}(a), \ldots, \sigma^{n-1}(a)\}.$$

The other orbits of $(a \quad b)\sigma$ coincide with orbits of σ.

So the number of orbits of $(a \quad b)\sigma$ exceeds the number of orbits of σ by one if $a \sim_\sigma b$. See Figure 12.3.

Figure 12.3: Composition with transposition $(a \quad b)$

Suppose a and b are in different orbits of σ. Let the orbit of a under σ be

$$\{a, \sigma(a), \sigma^2(a), \ldots, \sigma^{n-1}(a)\},$$

and the orbit of b

$$\{b, \sigma(b), \sigma^2(b), \ldots, \sigma^{m-1}(b)\}.$$

Then the orbit of a under $(a \quad b)\sigma$ is

$$\{a, \sigma(a), \ldots, \sigma_{n-1}(a), b, \sigma(b), \ldots, \sigma^{m-1}(b)\}.$$

The other orbits of $(a \quad b)\sigma$ coincide with orbits of σ.

So the number of orbits of $(a \quad b)\sigma$ is one less than the number of orbits of σ if not $a \sim_\sigma b$. See also for this case Figure 12.3. (This figure does not show whether a and b are in the same orbit.) □

12.21 Corollary. *Let σ be a permutation of a finite set A, and let τ be a transposition of A. Then*

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

Proof. From $\#(A) - \#(A_{\tau\sigma}) = \#(A) - (\#(A_\sigma) \pm 1)$ follows $\operatorname{sgn}(\tau\sigma) = -\operatorname{sgn}(\sigma)$. □

12.22 Proposition. *Let σ be a permutation of a finite set A, then σ is a product of $\#(A) - \#(A_\sigma)$ transpositions. (We consider the identical permutation as a product of 0 transpositions.)*

Proof. If σ has an orbit with more than one element, then choose two elements a_1 and b_1 in that orbit. The number of orbits of $(a_1 \quad b_1)\sigma$ then is 1 greater than the number of orbits of σ. Repeat this process. It stops as soon as all orbits have only 1 element. Then $(a_k \quad b_k)(a_{k-1} \quad b_{k-1}) \cdots (a_1 \quad b_1)\sigma = 1_A$, where $k = \#(A) - \#(A_\sigma)$. So the permutation σ is a product of k transpositions:

$$\sigma = (a_1 \quad b_1)(a_2 \quad b_2) \cdots (a_k \quad b_k).$$ □

From the proof it follows that $\#(A) - \#(A_\sigma)$ is in fact the least number of transpositions needed for writing σ as a product of transpositions.

Python

The following code returns for a permutation a list of a minimal number of transpositions having this permutation as product.

combinatorics.py

```
def transpositions(perm):
    trans = []
    newperm = perm[:]
    identity = list(range(1, len(perm) + 1))
    while newperm != identity:
        j = [i + 1 == newperm[i] for i in range(len(newperm))].\
        index(0)
        trans.append((j + 1, newperm[j]))
        newperm[newperm.index(j + 1)] = newperm[j]
        newperm[j] = j + 1
    return trans
```

```
>>> transpositions([4, 6, 8, 10, 13, 11, 9, 7, 1, 2, 5, 12, 3])
[(1, 4), (2, 6), (3, 8), (4, 10), (5, 13), (6, 11), (7, 9), (8, 9), (
9, 10), (10, 11), (11, 13)]
```

12.23 Theorem. *Let σ and τ be permutations of a finite set A. Then:*

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

Proof. Let $\sigma_1, \ldots, \sigma_n$ be transpositions such that $\sigma = \sigma_1 \cdots \sigma_n$. Then

$$\begin{aligned}\operatorname{sgn}(\sigma\tau) &= \operatorname{sgn}(\sigma_1 \cdots \sigma_n\tau) = -\operatorname{sgn}(\sigma_2 \cdots \sigma_n\tau) = \cdots \\ &= (-1)^n \operatorname{sgn}(\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau).\end{aligned}$$

□

12.24 Example. We write $\sigma = (1\quad 2\quad 3\quad 5)(4\quad 6\quad 7)$ as product of transpositions.

$$\begin{aligned}(1\quad 2)\sigma &= (2\quad 3\quad 5)(4\quad 6\quad 7)\\ (2\quad 3)(1\quad 2)\sigma &= (3\quad 5)(4\quad 6\quad 7)\\ (3\quad 5)(2\quad 3)(1\quad 2)\sigma &= (4\quad 6\quad 7)\\ (4\quad 6)(3\quad 5)(2\quad 3)(1\quad 2)\sigma &= (6\quad 7)\\ (6\quad 7)(4\quad 6)(3\quad 5)(2\quad 3)(1\quad 2)\sigma &= (1)\end{aligned}$$

So $\sigma = (1\quad 2)(2\quad 3)(3\quad 5)(4\quad 6)(6\quad 7)$.

12.25 Theorem. *Let $\sigma = \sigma_1 \cdots \sigma_m$ with $\sigma_1, \ldots, \sigma_m$ transpositions. Then*

$$m \text{ is even} \iff \sigma \text{ is an even permutation.}$$

Proof. $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma_1) \cdots \operatorname{sgn}(\sigma_m) = (-1)^m$. □

This means that an even permutation can only be written as a product of an even number of transpositions, and an odd permutation only as a product of an odd number of transpositions.

A 3-cycle is even. So any product of 3-cycles is even. In fact all even permutations are of this type:

12.26 Theorem. *Let A be a finite set with $\#(A) \geq 3$. Then every even permutation of A is a product of 3-cycles.*

Proof. Let σ be an even permutation. Then σ is a product of an even number of transpositions. For a composition of two transpositions we have the following cases

a) $(a_1 \quad a_2)(a_1 \quad a_2)$: this is equal to (1);
b) $(a_1 \quad a_2)(a_2 \quad a_3)$ with $a_3 \neq a_1$: this is equal to $(a_1 \quad a_3 \quad a_2)$;
c) $(a_1 \quad a_2)(a_3 \quad a_4)$ with a_1, a_2, a_3 and a_4 all different: this product is equal to $(a_1 \quad a_3 \quad a_2)(a_1 \quad a_3 \quad a_4)$.

Using this we see that the product of an even number of transpositions is also a product of 3-cycles. □

Python

The sign of a permutation equals -1 to the power the number of elements minus the number of orbits.

combinatorics.py

```
def sign(perm):
    return (-1)**(len(perm) - len(cycledecomposition(perm)))
```

```
>>> sign([12, 1, 13, 4, 5, 9, 7, 6, 3, 2, 8, 10, 11])
1
```

Sam Loyd's 14-15-puzzle

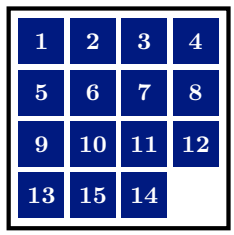
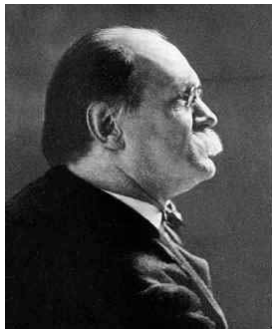


Figure 12.4: 14-15-puzzle

The sliding puzzle of example 3.3 is introduced in 1878 by Sam **Loyd**. The tiles were placed in the right order except for the last two. That is why the puzzle is known as the 14-15-puzzle. Loyd offered a prize of \$1000 for the solution, knowing that the puzzle has no solution. A worldwide madness was the result. Companies had to forbid their employees trying to solve it during office hours.

Samuel Loyd (Philadelphia 1841 – New York 1911)



Samuel Loyd was active in recreational mathematics. He designed many puzzles, some of which became known all over the world. Another well-known puzzle is the 'Get off the Earth'-puzzle.

Coding positions and moves

We have a set Tiles of 16 tiles, among them one 'virtual' tile, and the set Places of the 16 possible places for the tiles. A position of the puzzle is a bijection $f\colon \text{Tiles} \to \text{Places}$. For coding the positions we choose a numbering for the tiles and for the places:

$$\text{tile}\colon \underline{16} \to \text{Tiles} \qquad \text{and} \qquad \text{place}\colon \underline{16} \to \text{Places}.$$

The tiles are numbered in such a way that tile(16) is the virtual tile. In the puzzle the tile numbers 1 up to 15 are displayed on the tiles.

A position f now corresponds to a permutation of $\underline{16}$:

$$\begin{array}{ccc} \underline{16} & \xrightarrow{\ \sigma\ } & \underline{16} \\ {\scriptstyle \text{tile}}\downarrow & & \downarrow{\scriptstyle \text{place}} \\ \text{Tiles} & \xrightarrow[\ f\ ]{} & \text{Places} \end{array}$$

A position $f\colon \text{Tiles} \to \text{Places}$ corresponds to a permutation σ of $\underline{16}$ if

$$\sigma(i) = j \iff f(\text{tile}(i)) = \text{place(j)}.$$

Let's denote the position having tile i on place $\sigma(i)$ by $[\sigma]$.

The puzzle can be turned into a graph. The vertices are the positions $[\sigma]$, there are 16! of them. Next we describe the edges. If τ is a permutation of $\underline{16}$, then in position $[\tau\sigma]$ tile i is on place $\tau(\sigma(i))$: the place being at first j, changes into $\tau(j)$. Replacing $[\sigma]$ by $[\tau\sigma]$ means changing the places of the tiles according to τ. Thus a position is transformed into a new position. It depends on $\sigma(16)$ which moves are possible when starting from $[\sigma]$. There are two, three or four moves possible depending on the place of tile 16. Each of these moves corresponds to a transposition of 16 with another number.

The positions that can be reached

A succession of moves starting from position $[\sigma]$ corresponds to a succession of multiplications (= compositions) on the left by transpositions. If 'tile' 16 is after these moves back in place $\sigma(16)$, then there have been an even number of vertical moves and an even number of horizontal moves. Together this corresponds to an even number of transpositions. If $[\sigma']$ is the result of the succession of moves, then σ and σ' have the same sign. So in particular the 14-15-puzzle is not solvable: the initial position corresponds to an odd permutation and the final position to an even one.

For the unsolvability of the 14-15-puzzle there was no need to number the places explicitly. By now it is convenient to give a numbering. The place numbers are indicated in Figure 12.5. In this figure tile(i) is on place(i). So with this numbering of the places Figure 12.5 displays the position $[(1)]$. If a position $[\tau]$ with $\tau(16) = 16$ is reached from this by a succession of moves, τ is an even permutation.

10	9	6	5
11	8	7	4
12	15	1	3
13	14	2	16

Figure 12.5: numbering of places

We will show that conversely all positions $[\tau]$ with τ even and $\tau(16) = 16$ can be reached. We look at the totality of successions of moves under which tile(16) returns to place(16). Two of such successions after one another is again such a succession. If positions $[\tau_1]$ and $[\tau_2]$ can be reached, then also position $[\tau_1\tau_2]$. If tile(16) is moved along the places 2, 1, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3 and finally returns in place 16, then the new position is $[\sigma]$, where

$$\sigma = (1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12 \quad 13 \quad 14 \quad 15).$$

After being moved along 2, 1, 3, 16 the position corresponds to the permutation

$$\rho = (1 \quad 2 \quad 3).$$

Since for $k \in \mathbb{N}_{12}$ we have

$$\sigma^k \rho \sigma^{-k} = (k+1 \quad k+2 \quad k+3),$$

it follows from lemma 12.27 below that each position $[\tau]$ with τ even can be reached.

12.27 Lemma. *Let $\in \mathbb{N}$ with $n \geq 3$. Then every even permutation of $\underline{n}$ is a product of 3-cycles of type $(k \quad k+1 \quad k+2)$ with $k \in \mathbb{N}_{n-2}$.*

Proof. Let σ be an even permutation of $\underline{n}$. A permutation can be denoted with the images of the elements;

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}$$

Right multiplication with a 3-cycle $(k \;\; k+1 \;\; k+2)$ yields

$$\sigma \cdot (k \quad k+1 \quad k+2) = \begin{pmatrix} 1 & \dots & k-1 & k & k+1 & k+2 & k+3 & \dots & n \\ \sigma(1) & \cdots & \sigma(k-1) & \sigma(k+1) & \sigma(k+2) & \sigma(k) & \sigma(k+3) & \cdots & \sigma(n) \end{pmatrix}.$$

In the bottom row $\sigma(k+1)$ is moved one place to the left. By multiplications with 3-cycles on the right it can be achieved that 1 is on the left, 2 on the second place and so on. This is repeated until $n-2$ is on place $n-2$. Since σ is even all these newly obtained permutations are even as well. So it can not be the case that 1 up to $n-2$ remain in place while $n-1$ and n do not. So we have 3-cycles of the indicated type, say $\tau_1, \dots, \tau_m$ such that $\sigma\tau_1\tau_2 \dots \tau_m = (1)$, that is $\sigma = \tau_m^2 \cdots \tau_1^2$. □

Exercises

1. We define polynomials $x^{\overline{n}}$ as follows

$$x^{\overline{n}} = x(x+1)(x+2)\cdots(x+n-1).$$

More precisely:

$$\begin{cases} x^{\overline{0}} = 1, \\ x^{\overline{n+1}} = x^{\overline{n}} \cdot (x+n) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Prove that for all $n \in \mathbb{N}$:

$$x^{\overline{n}} = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix} x^k.$$

2. Write the following permutations as a product of transpositions:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 2 & 5 & 1 & 3 & 4 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 1 & 4 & 5 & 6 & 2 \end{pmatrix}.$$

3. Which permutations of $\underline{4}$ are even?

4. Let A be a finite set with at least two elements. Prove that the number of even permutations of A equals $\frac{1}{2}n!$.

5. Prove that

$$\sum_{k=0}^{n} (-1)^k \begin{bmatrix} n \\ k \end{bmatrix} = \begin{cases} 1 & \text{if } n = 0 \\ -1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases}$$

6. Let $n \in \mathbb{N}$ with $n > 1$. Determine the sign of an n-cycle. Write the n-cycle $(1 \quad 2 \quad \cdots \quad n)$ as product of a minimal number of transpositions.

7. Let $a_1, a_2, \ldots, a_n$ be different elements of a set A and let σ be a permutation of A. Show that

$$\sigma(a_1 \quad a_2 \quad \cdots \quad a_n)\sigma^{-1} = (\sigma(a_1) \quad \sigma(a_2) \quad \cdots \quad \sigma(a_n)).$$

8. (i) Show that $\begin{bmatrix} n+1 \\ 1 \end{bmatrix} = n!$ for all $n \in \mathbb{N}$.

   (ii) Show that $\begin{bmatrix} n \\ n-1 \end{bmatrix}$ is a triangular number for all $n \in \mathbb{N}^+$.

   (iii) Determine $\displaystyle\sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}$.

9. Let $k \in \mathbb{N}$. Is the sequence $\left(\begin{bmatrix} n+k \\ n \end{bmatrix} \right)$ a polynomial sequence?

10. Let σ be a permutation of $A = \underline{15}$ with $\sigma^5 = 1_A$ and $\sigma(i) \neq i$ for all $i \in A$.
    (i) Show that σ is a product of three disjoint 5-cycles.
    (ii) Determine the sign of σ.
    (iii) Show that there are transpositions τ_1 and τ_2 such that $\tau_2\tau_1\sigma$ is a 15-cycle.

11. Let $n \in \mathbb{N}^+$. How many of the permutations of $\underline{n}$ do have exactly one orbit?

12. Let σ and τ be permutations of a finite set A. Show that the number of orbits of σ in A is equal to the number of orbits of $\tau\sigma\tau^{-1}$.

13. Let σ and τ be permutations of a finite set A. Show that $\sigma\tau\sigma^{-1}\tau^{-1}$ is an even permutation.

14. Let $V_1 = \{1, 2, 3, \ldots, 100\}$, $V_2 = \{101, 102, 103, \ldots, 200\}$ and $V = V_1 \cup V_2$. Let σ be a permutation of V. Show that

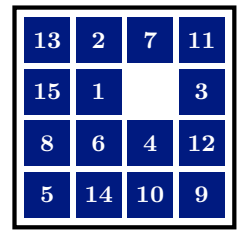
$$\#\{\, i \in V_1 \mid \sigma(i) \in V_2 \,\} = \#\{\, j \in V_2 \mid \sigma(j) \in V_1 \,\}.$$

15. How many permutations σ of $\underline{100}$ are there with $\sigma(a)$ even for all $a \leq 50$?

16. We denote the numbers in $\mathbb{N}_{1000000}$ using 6 digits: so 000000, 000001, 000002 up to 999999. The permutation σ of $\mathbb{N}_{1000000}$ maps a number to the number obtained by shifting the digits in this notation one place to the right and putting the left most digit in front, so for example $\sigma(123456) = 612345$ and $\sigma(000562) = 200056$.
    (i) Describe the orbit of σ which contains the element 264264.
    (ii) Show that the orbits of this permutation have 1, 2, 3 or 6 elements. Give for each number of elements an example.
    (iii) How many of the orbits of σ have 6 elements?
    (iv) Determine $\operatorname{sgn}(\sigma)$.

17. Let σ and τ be permutations of a finite set.
    (i) Prove that σ and $\tau\sigma\tau^{-1}$ have the same number of orbits.
    (ii) Prove that $\sigma\tau$ and $\tau\sigma$ have the same number of orbits.

18. Is the sliding puzzle with this starting position solvable?



13 Modular Arithmetic

Arithmetic modulo $m \in \mathbb{N}^+$ is arithmetic with integers, where numbers are seen as 'the same' when they differ by a multiple of m. In fact this is an equivalence relation and the arithmetic is done with equivalence classes. The objects of the arithmetic being clear, the arithmetic itself still has to be described, that is addition and multiplication have to be defined and these operations have to satisfy the usual rules. Arithmetic modulo m is arithmetic in a set with m elements. In this chapter and the next we study the structure of the ring $\mathbb{Z}/m$ thus obtained. In chapter 15 applications of modular arithmetic will be given: prime tests, the factorization of integers, the RSA-code (a cryptographic application).

13.1 Residue Classes Modulo m

13.1 Definition. Let $m \in \mathbb{N}^+$. For $a, b \in \mathbb{Z}$ we define

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

If $a \equiv b \pmod{m}$ we say that *a is congruent with b modulo m*. "Congruence modulo m" is a relation in $\mathbb{Z}$. For each $m \in \mathbb{N}^+$ we have a congruence.

13.2 Proposition. *Congruence modulo m is a equivalence relation.*

Proof. For every $a \in \mathbb{Z}$ we have $m \mid a - a$, and so the relation is reflexive. From $m \mid a - b$ follows $m \mid b - a$, which means that the relation is symmetric. Transitivity: if $m \mid a - b$ and $m \mid b - c$, then by proposition 9.24 $m \mid (a - b) + (b - c) = a - c$. □

13.3 Definition. Let $m \in \mathbb{N}^+$. An equivalence class of the congruence modulo m is called a *residue class* modulo m. The set of residue classes modulo m is denoted by $\mathbb{Z}/m$. The residue class of a modulo m is denoted by $[a]_m$ or as $[a]$ or even $\overline{a}$ when it is clear which m is used.

13.4 Proposition. *Let $m \in \mathbb{N}^+$. Then $\mathbb{N}_m$ is a system of representatives of $\mathbb{Z}/m$. In particular $\#(\mathbb{Z}/m) = m$.*

Proof. Let $a \in \mathbb{Z}$. Division with remainder of a by m gives $a = qm + r$ with $q \in \mathbb{Z}$ and $r \in \mathbb{N}_m$. From $a - r = qm$ follows $m \mid a - r$, that is $a \equiv r \pmod{m}$. So $r \in \overline{a}$ and in each residue class there is an element of $\mathbb{N}_m$, even a unique element: if also $s \in \overline{r}$ with $s \in \mathbb{N}_m$, then $m \mid r - s$ while $-m < r - s < m$. So $r - s = 0$, that is $r = s$. □

Division with remainder by m is a surjective map $q_m\colon \mathbb{Z} \to \mathbb{N}_m$. The induced partition of $\mathbb{Z}$ is $\mathbb{Z}/m$. Numbers are congruent modulo m if the remainders after dividing by m are equal. The map $q_m\colon \mathbb{Z} \to \mathbb{N}_m$ induces a bijection $\mathbb{Z}/m \to \mathbb{N}_m$.

So we have $\mathbb{Z}/m = \{\overline{0}, \overline{1}, \dots, \overline{m-1}\}$ and the elements $\overline{0}, \overline{1}, \dots, \overline{m-1}$ are different. We describe the residue class $\overline{k}$. By definition

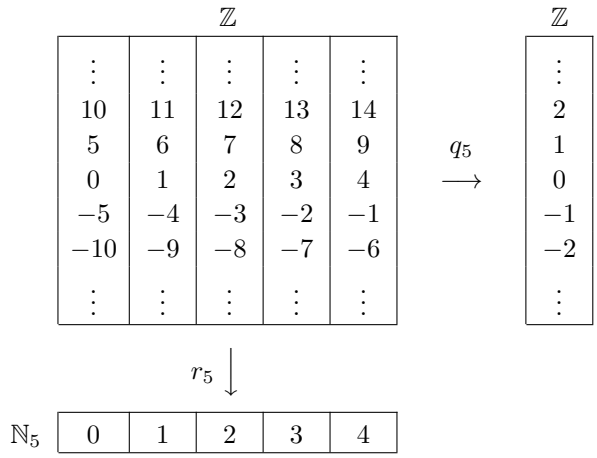
$$\overline{k} = \{\, a \in \mathbb{Z} \mid a \equiv k \pmod{m} \,\},$$

and $a \equiv k \pmod{m}$ means that $m \mid a - k$, which in turn means that there is a $t \in \mathbb{Z}$ such that $a - k = tm$, that is $a = k + tm$. The residue class $\overline{k}$ is the set of all multiples of m plus k. It is the set of the numbers

$$\dots,\ k-2m,\ k-m,\ k,\ k+m,\ k+2m,\ \dots,\ k+tm,\ k+(t+1)m,\ \dots.$$

The classes $\overline{a}$ are the orbits of the permutation $x \mapsto x + m$ of $\mathbb{Z}$.

Below the partition of $\mathbb{Z}$ into 5 residue classes modulo 5 is displayed:



13.2 The Ring $\mathbb{Z}/m$

In this section m is a fixed nonzero natural number. We will define addition in the set $\mathbb{Z}/m$ of residue classes modulo m. The addition will be induced by the addition of integers.

13.5 Lemma. *Let $a, a', b, b' \in \mathbb{Z}$ satisfy $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$ (in $\mathbb{Z}/m$). Then $\overline{a+b} = \overline{a'+b'}$.*

Proof. Since $m \mid a-a'$ and $m \mid b-b'$, we have $m \mid a-a'+b-b' = (a+b)-(a'+b')$, that is $\overline{a+b} = \overline{a'+b'}$. □

13.6 Definition. Let a and b be integers. We define the *sum* $\overline{a}+\overline{b}$ of $\overline{a}$ and $\overline{b}$:

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

The residue classes $\overline{a}$ and $\overline{b}$ are given by their representatives a and b. That could have been others, say a' and b'. From lemma 13.5 it follows that the result is independent of the choice of the representatives. That is why this sum of residue classes is well-defined.

13.7 Proposition. *The set $\mathbb{Z}/m$ is together with the addition defined above is an abelian group.*

Proof. The proof is simple, but since this proposition is of some importance, details will be given. We have to verify the abelian group axioms: associativity, commutativity, existence of a zero element, existence of opposites.

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

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

Zero element: $\overline{0}$ is the zero element: $\overline{a}+\overline{0}=\overline{a+0}=\overline{a}$.

Opposite: $\overline{-a}$ is the opposite of $\overline{a}$: $\overline{a}+\overline{-a}=\overline{a+(-a)}=\overline{0}$. □

So the fact that $\mathbb{Z}/m$ under this addition is an abelian group, is a direct consequence of $\mathbb{Z}$ being an abelian group under addition. The advantage of this approach with residue classes is that this proof is straightforward. If we had chosen for the set $\mathbb{N}_m$, then addition could have been defined using division with remainder: the sum of a and b then is the remainder of $a+b$ after division by m. To prove that the structure thus defined is an abelian group is much more elaborate: for associativity for example many cases have to be distinguished.

The additive structure of $\mathbb{Z}/m$ as defined here is quite simple, which will become clear when making up an addition table:

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\cdots$	$\overline{m-2}$	$\overline{m-1}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\cdots$	$\overline{m-2}$	$\overline{m-1}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\cdots$	$\overline{m-1}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\cdots$	$\overline{0}$	$\overline{1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$\overline{m-2}$	$\overline{m-2}$	$\overline{m-1}$	$\overline{0}$	$\overline{1}$	$\cdots$	$\overline{m-4}$	$\overline{m-3}$
$\overline{m-1}$	$\overline{m-1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\cdots$	$\overline{m-3}$	$\overline{m-2}$

The definition of multiplication in $\mathbb{Z}/m$ is similar to the definition of addition.

13.8 Lemma. *Let $a, a', b, b' \in \mathbb{Z}$ satisfy $\overline{a} = \overline{a'}$ and $\overline{b} = \overline{b'}$ (in $\mathbb{Z}/m$). Then: $\overline{ab} = \overline{a'b'}$.*

Proof. Since $m \mid a - a'$ and $m \mid b - b'$, we also have $m \mid (a - a')b + a'(b - b') = ab - a'b'$, that is $\overline{ab} = \overline{a'b'}$. □

The definition of multiplication is based on this lemma.

13.9 Definition. Let a and b be integers. We define the *product* $\overline{a} \cdot \overline{b}$ of $\overline{a}$ and $\overline{b}$:

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

The $\cdot$ is there for clarity and is often omitted.

13.10 Theorem. *$(\mathbb{Z}/m, +, \cdot)$ is a commutative ring.*

Proof.

$\mathbb{Z}/m$ is an abelian group under the addition: This is proposition 13.7.

Associativity of the multiplication: $(\overline{a} \cdot \overline{b}) \cdot \overline{c} = \overline{ab} \cdot \overline{c} = \overline{(ab)c} = \overline{a(bc)} = \overline{a} \cdot (\overline{bc}) = \overline{a} \cdot (\overline{b} \cdot \overline{c})$.

Commutativity: $\overline{a} \cdot \overline{b} = \overline{ab} = \overline{ba} = \overline{b} \cdot \overline{a}$.

Unity element: $\overline{1}$ is the unity element: $\overline{1} \cdot \overline{a} = \overline{1a} = \overline{a}$.

Distributivity: $\overline{a} \cdot (\overline{b} + \overline{c}) = \overline{a} \cdot \overline{b + c} = \overline{a(b + c)} = \overline{ab + ac} = \overline{ab} + \overline{ac} = \overline{a} \cdot \overline{b} + \overline{a} \cdot \overline{c}$. Similarly $(\overline{a} + \overline{b}) \cdot \overline{c} = \overline{a} \cdot \overline{c} + \overline{b} \cdot \overline{c}$. □

So the fact that $\mathbb{Z}/m$ is a commutative ring is an easy consequence of $\mathbb{Z}$ being a commutative ring.

The multiplication table of $\mathbb{Z}/m$ is more complicated than its addition table. The multiplication tables for $2 \leq m \leq 6$:

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

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

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

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{1}$	$\overline{3}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{1}$	$\overline{4}$	$\overline{2}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

$\cdot$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{0}$	$\overline{2}$	$\overline{4}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{0}$	$\overline{3}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{2}$	$\overline{0}$	$\overline{4}$	$\overline{2}$
$\overline{5}$	$\overline{0}$	$\overline{5}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

Python

For representing the elements of $\mathbb{Z}/m$ usually the system $\mathbb{N}_m$ of representatives is used. Addition and multiplication is done for integers. Results are replaced by their remainder after division by m. Thus the size of the numbers remains restricted whatever the number of operations. We add the code to the module `arithmetics.py`.

arithmetics.py

```
def modsum(x, y, m):
    return (x + y) % m

def modprod(x, y, m):
    return (x * y) % m
```

```
>>> modsum(53679298709, 456297098, 2345)
2162
>>> modprod(53679298709, 456297098, 2345)
712
```

13.3 Exponentiation in $\mathbb{Z}/m$

Exponentiation in a ring is repeated multiplication:

$$\begin{cases} a^0 = 1 \\ a^{n+1} = a^n \cdot a \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

The computation of a^n can be reduced to two operations: squaring and multiplication by a. To understand how this works it is convenient to use the binary notation for the exponent n. Squaring maps a^k to a^{2k}. In the binary notation for k this means that a 0 is added on the right. A 1 added on the right results in squaring followed by multiplication with a: then a^k maps to a^{2k+1}. The binary notation of n tells you exactly what to do to reach a^n using these operations.

13.11 Example. We compute 7^{44} in $\mathbb{Z}$. First determine the binary notation of 44. It is $[1,0,1,1,0,0]_2$, or 101100 for short. Here the binary notation is in red. We have:

$$\begin{aligned}
7^{1} &= 7\\
7^{10} &= 7^2 = 49\\
7^{101} &= 49^2 \cdot 7 = 2401 \cdot 7 = 16807\\
7^{1011} &= 16807^2 \cdot 7 = 282475249 \cdot 7 = 1977326743\\
7^{10110} &= 1977326743^2 = 3909821048582988049\\
7^{101100} &= 3909821048582988049^2 = 15286700631942576193765185769276826401.
\end{aligned}$$

When doing arithmetic modulo m intermediate results can be replaced by their remainders after dividing by m.

13.12 Example. We now compute $\overline{7}^{44}$ in $\mathbb{Z}/33$. With the exponents on the left hand side in binary notation and $\equiv$ to emphasize that the numbers are representatives:

$$\begin{aligned}
7^{1} &= 7\\
7^{10} &= 7^2 = 49 \equiv 16\\
7^{101} &\equiv 16^2 \cdot 7 = 256 \cdot 7 \equiv 25 \cdot 7 = 175 \equiv 10\\
7^{1011} &\equiv 10^2 \cdot 7 = 100 \cdot 7 \equiv 1 \cdot 7 = 7\\
7^{10110} &\equiv 7^2 = 49 \equiv 16\\
7^{101100} &\equiv 16^2 = 256 \equiv 25.
\end{aligned}$$

So in $\mathbb{Z}/33$ we have $\overline{7}^{44} = \overline{25}$. Note that the numbers are not as large as in the previous example. We could have computed 7^{44} first and finally determine the remainder after division by 33. But certainly when computing something like $\overline{7}^{543678299982762}$ this will not work.

Python

First compute the binary representation of a natural number:

arithmetics.py

```
def binary(x):
    bin = []
    while x != 0:
        d = divmod(x, 2)
        x, bin = d[0], [d[1]] + bin
    return bin
```

Code for exponentiation in $\mathbb{Z}/m$:

arithmetics.py

```
def modpower(x, y, m):
    bin = binary(y)
    result = 1
    for i in bin:
        result = modprod(result, result, m)
        if i == 1:
            result = modprod(result, x, m)
    return result
```

The function `modpower` calls the function `binary`.

```
>>> binary(24569209)
[1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0,
 0, 1]
>>> modpower(20029727, 24569209, 356582)
114547
```

In Python there is standard the function `pow(x,y,z)` which has the same effect as `modpower`. That is why we will use `pow` from now on.

```
>>> pow(20029727, 24569209, 356582)
114547
```

As in example 13.12 we see that in $\mathbb{Z}/33$ one has $\overline{7}^{10} = \overline{1}$ (the exponent 10 is here decimal, binary it is 1010). In Section 13.5 we will have a closer look at such phenomena. From Euler's theorem in that section it follows that in this case $\overline{7}^{20} = \overline{1}$. In Section 13.6 we will understand why $\overline{7}^{10} = \overline{1}$ without doing a lot of computation.

13.4 Invertible Elements Modulo m

This section is about the multiplicative structure of the ring $\mathbb{Z}/m$ and in particular the invertible elements, in other words it is about the group $(\mathbb{Z}/m)^*$ of the invertible elements in $\mathbb{Z}/m$. See notation 9.9.

In the tables in section 13.2 we see that $(\mathbb{Z}/2)^* = \{\overline{1}\}$, $(\mathbb{Z}/3)^* = \{\overline{1}, \overline{2}\}$, $(\mathbb{Z}/4)^* = \{\overline{1}, \overline{3}\}$, $(\mathbb{Z}/5)^* = \{\overline{1}, \overline{2}, \overline{3}, \overline{4}\}$ and $(\mathbb{Z}/6)^* = \{\overline{1}, \overline{5}\}$. So three of these five rings are fields. We have fields with 2, 3 and 5 elements. We will see that $\mathbb{Z}/p$ is a field for every prime number p.

13.13 Proposition. *Let a be an integer. Then:*

$$\overline{a} \in (\mathbb{Z}/m)^* \iff \gcd(a, m) = 1.$$

Proof. Equivalent are:

$\overline{a} \in (\mathbb{Z}/m)^*$.

There is an $x \in \mathbb{Z}$ such that $\overline{x} \cdot \overline{a} = \overline{1}$.

There is an $x \in \mathbb{Z}$ such that $\overline{xa} = \overline{1}$.

There is an $x \in \mathbb{Z}$ such that $xa \equiv 1 \pmod{m}$.

There is an $x \in \mathbb{Z}$ such that $m \mid xa - 1$.

There are $x, y \in \mathbb{Z}$ such that $ym = xa - 1$.

There are $x, y \in \mathbb{Z}$ such that $xa + (-y)m = 1$.

$\gcd(a, m) = 1$. □

Algorithm

For determining whether an $\overline{a} \in \mathbb{Z}/m$ is invertible it suffices to verify $\gcd(a, m) = 1$ and that can be done very fast by Euclid's algorithm (section 9.4). The extended version of Euclid's algorithm described on page 156 can be used to determine the inverse of an $\overline{a} \in (\mathbb{Z}/m)^*$.

Python

We add a function which determines the inverse in modular arithmetic. It calls the extended Euclidean algorithm.

arithmetics.py

```
def modinv(x, m):
    return euclid(x, m)[0] % m
```

```
>>> gcd(244552, 177277)
1
>>> modinv(244552, 177277)
18148
```

The ring $\mathbb{Z}/m$ is a field if all elements $\neq \overline{0}$ are invertible:

13.14 Theorem. *$\mathbb{Z}/m$ is a field if and only if m is a prime number.*

Proof. Equivalent are:

$\mathbb{Z}/m$ is a field.

$\overline{a} \in (\mathbb{Z}/m)^*$ for all $a \in \mathbb{Z}$ with $\overline{a} \neq \overline{0}$.

$\gcd(a, m) = 1$ for all $a \in \mathbb{Z}$ with $m \nmid a$.

$\gcd(a, m) = 1$ or $\gcd(a, m) = m$ for all $a \in \mathbb{Z}$.

m is a prime number. □

13.15 Notation. So for every prime number p there is a finite field: the field $\mathbb{Z}/p$ with p elements. Fields are important in mathematics. That is why for these fields there is a special notation: $\mathbb{F}_p$.

> The F in this notation comes from 'field', the English name for this notion. Other finite fields than these exist. The number of elements of a finite field can only be a power of a prime number. In fact there is a complete classification: up to isomorphy there is a unique field for each prime power. We will not prove this here. See section 20.2 for a construction of finite fields with the square of a prime number as their number of elements.

Equations over $\mathbb{F}_p$

Let p be a prime number. Because $\mathbb{F}_p$ is a field, all of section 9.2 on the solution of polynomial equations is applicable to equations over $\mathbb{F}_p$.

13.16 Example. We solve in $\mathbb{F}_{17}$ the linear equation

$$\overline{3}x + \overline{10} = \overline{0}.$$

First we multiply with the inverse of $\overline{3}$, which is $\overline{6}$:

$$\overline{6} \cdot \overline{3}x + \overline{6} \cdot \overline{10} = x + \overline{9} = \overline{0}.$$

So $x = -\overline{9} = \overline{-9} = \overline{8}$.

13.17 Example. We solve in $\mathbb{F}_{17}$ the quadratic equation

$$\overline{3}x^2 + \overline{16}x + \overline{5} = \overline{0}.$$

First we multiply with the inverse of $\overline{3}$, so with $\overline{6}$:

$$x^2 + \overline{11}x + \overline{13} = \overline{0}.$$

Next we 'complete the square' (note that $\overline{11} = \overline{28} = \overline{2} \cdot \overline{14}$):

$$(x + \overline{14})^2 - \overline{14}^2 + \overline{13} = \overline{0},$$

So

$$(x + \overline{14})^2 = \overline{14}^2 - \overline{13} = \overline{9} - \overline{13} = \overline{13} = \overline{8}^2,$$

and so

$$(x + \overline{14})^2 - \overline{8}^2 = (x + \overline{6})(x + \overline{22}) = \overline{0}.$$

The solutions are $x = \overline{11}$ and $x = \overline{12}$.

> For $p \neq 2$ the solution of a quadratic equation in $\mathbb{F}_p$ comes down to extracting a square root in $\mathbb{F}_p$. So the problem is whether an element of this finite field is a square, and if so what is its square root? For the first there is a solution as fast as the Euclidean algorithm. Extracting square roots then still is a problem. We go into this in chapter 14.

13.18 Example. The following equation over $\mathbb{F}_{17}$

$$x^4 - \overline{1} = \overline{0} \tag{13.1}$$

has the solutions $\overline{1}$, $-\overline{1}$, $\overline{4}$ and $-\overline{4}$. Since the degree of this equation is 4, by theorem 9.21 these are all solutions.

In the same field $\overline{6}$ is a solution of

$$x^4 - \overline{4} = \overline{0}.$$

Using the solutions of equation (13.1) three more solutions of this equation are found: $-\overline{6}$, $\overline{4} \cdot \overline{6}$ $(= \overline{7})$ and $-\overline{7}$. In the last paragraph of this chapter this kind of computation is described in general.

13.5 Euler's Theorem

In this section we study the abelian group $(\mathbb{Z}/m)^*$ of invertible elements in $\mathbb{Z}/m$. We have already seen that for $a \in \mathbb{Z}$:

$$\overline{a} \in (\mathbb{Z}/m)^* \iff \gcd(a, m) = 1.$$

So theorem 10.42 implies:

13.19 Proposition. *Let $m \in \mathbb{N}^+$. Then $\#((\mathbb{Z}/m)^*) = \varphi(m) = m \cdot \prod_{p|m}(1 - \frac{1}{p})$.* □

In the next section another proof for this formula for the totient will be given.

Let $\overline{a} \in (\mathbb{Z}/m)^*$. Multiplication by $\overline{a}$ is a permutation of $(\mathbb{Z}/m)^*$, multiplication by $\overline{a}^{-1}$ being the inverse permutation. We denote the multiplication by $\overline{a}$ as σ_a:

$$\sigma_a \colon (\mathbb{Z}/m)^* \to (\mathbb{Z}/m)^*, \quad \overline{c} \mapsto \overline{a} \cdot \overline{c}\,(= \overline{ac}).$$

If the orbit of a $\overline{c}$ under the permutation σ_a has k elements, then this orbit is the subset

$$\{\overline{c}, \overline{ac}, \overline{a^2c}, \ldots, \overline{a^{k-1}c}\},$$

where k is the least in $\mathbb{N}^+$ such that $\overline{a^k c} = \overline{c}$, that is $\overline{a^k} = \overline{1}$. So the size of the orbit does not depend on $\overline{c}$: all orbits are equal in size! If there are r orbits and each of them has k elements, then $rk = \varphi(m)$.

13.20 Definition. Let $m \in \mathbb{N}^+$ and $a \in \mathbb{N}$ with $\gcd(a, m) = 1$. The least $k \in \mathbb{N}^+$ with $\overline{a}^k = \overline{1}$ is called the *order* of a modulo m. Notation: $\mathrm{o}_m(a) = k$.

> More generally this notion of *order* is applicable to an element of a group. The order might be infinite (if such least k is not there, the orbit is infinite). For example: the element 1 in the additive group $\mathbb{Z}$. The orbit of $a \mapsto a + 1$ is the entire set $\mathbb{Z}$. Also the permutations of the set $\underline{n}$ (with $n \in \mathbb{N}^+$) form a group: the operation is the composition of permutations. A k-cycle is an example of a permutation of order k. The order of a product of disjoint cycles is the least common multiple of the orders of these cycles.

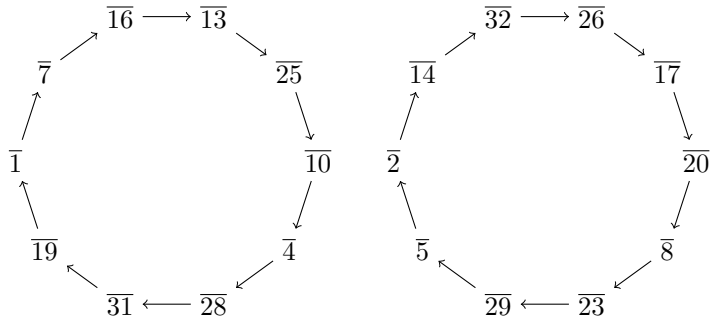


Figure 13.1: The permutation σ_7 of $\mathbb{Z}/33^*$.

13.21 Example. In example 13.12 we computed $\overline{7}^{44}$ in the ring $\mathbb{Z}/33$. Since $\gcd(7, 33) = 1$, the element $\overline{7}$ is invertible in $\mathbb{Z}/33$, that is $\overline{7} \in (\mathbb{Z}/33)^*$. The number of elements of the group $(\mathbb{Z}/33)^*$ is $\varphi(33) = \varphi(3)\varphi(11) = 2 \cdot 10 = 20$. Figure 13.1 is a picture of the permutation σ_7 of the set $(\mathbb{Z}/33)^*$. It is a product of two disjoint cycles, each of length 10. The powers of $\overline{7}$ are in the orbit on the left: $\overline{7}^0 = \overline{1}, \overline{7}^1 = \overline{7}, \overline{7}^2 = \overline{16}, \ldots$. Starting in $\overline{1}$, after 40 steps the orbit is completed 4 times and after an extra 4 steps you arrive at $\overline{25}$. The orbit on the right is obtained from the orbit on the left by multiplying each element by $\overline{2}$ (or any other element of the orbit on the right). Multiplication by $\overline{2}$ has the effect of swapping the two orbits.

The order is a divisor of the totient. This leads directly to the following theorem.

13.22 Theorem (Euler). *Let $m \in \mathbb{N}^+$. Then for all $a \in \mathbb{N}$ such that $\gcd(a, m) = 1$*

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

Proof. Let r be the number of orbits of σ_a. Then $\varphi(m) = \mathrm{o}_m(a) \cdot r$ and so

$$\overline{a}^{\varphi(m)} = \overline{a}^{\mathrm{o}(a)_m \cdot r} = (\overline{a}^{\mathrm{o}(a)_m})^r = \overline{1}^r = \overline{1}. \qquad \square$$

The special case $m = p$ with p a prime number yields:

13.23 Corollary (Fermat's Little Theorem). *Let p be a prime number. Then*

$$a^{p-1} \equiv 1 \pmod{p} \qquad \text{for all } a \in \mathbb{Z} \text{ with } p \nmid a,$$

and so also

$$a^p \equiv a \pmod{p} \qquad \text{for all } a \in \mathbb{Z}. \qquad \square$$

So the equation $x^p - x = \overline{0}$ over $\mathbb{F}_p$ has p solutions: all elements of the field satisfy the equation.

If we know the prime factorization of m, then we can easily determine $\varphi(m)$. As a result powers of invertible elements of $\mathbb{Z}/m$ are easily computed using Euler's theorem.

13.24 Example. We compute $\overline{2}^{1000}$ in $\mathbb{Z}/45$. We have $\varphi(45) = \varphi(9)\varphi(5) = 6 \cdot 4 = 24$. Since $\gcd(2, 45) = 1$ we have by Euler's theorem: $\overline{2}^{24} = \overline{1}$. Division with remainder yields $1000 = 41 \cdot 24 + 16$, and so $\overline{2}^{1000} = \overline{2}^{16}$. And $\overline{2}^{16}$ is easily computed, for example: $\overline{2}^{16} = \overline{4}^{8} = \overline{16}^{4} = \overline{31}^{2} = \overline{16}$, or $\overline{2}^{16} = \overline{2}^{12} \cdot \overline{2}^{4} = \overline{2}^{4}$, because $\overline{2}^{12} = \overline{1}$.

> Note that $\varphi(m)$ is not necessarily the least number k in $\mathbb{N}^+$ such that $a^k \equiv 1 \pmod{m}$ for all a with $\gcd(a, m) = 1$. For example $\varphi(8) = 4$, while $\overline{a}^2 \equiv \overline{1}$ for all $\overline{a} \in (\mathbb{Z}/8)^*$ $(= \{\overline{1}, \overline{3}, -\overline{3}, -\overline{1}\})$.

13.6 The Chinese Remainder Theorem

For $m, n \in \mathbb{N}^+$ with $\gcd(m, n) = 1$ arithmetic modulo mn comes down to arithmetic simultaneously modulo m and modulo n. That is what the Chinese Remainder Theorem is about.

13.25 Chinese Remainder Theorem. *Let $m, n \in \mathbb{N}^+$ satisfy $\gcd(m, n) = 1$. Then to every pair a, b of integers there is an integer c such that $c \equiv a \pmod{m}$ and $c \equiv b \pmod{n}$. If also d is an integer such that $d \equiv a \pmod{m}$ and $d \equiv b \pmod{n}$, then $d \equiv c \pmod{mn}$.*

Proof. There are $x, y \in \mathbb{Z}$ such that $xm + yn = 1$. Take $c = ayn + bxm$. Then $c \equiv ayn \pmod{m}$ and so, since $yn \equiv 1 \pmod{m}$ and $c \equiv a \pmod{m}$. Similarly $c \equiv b \pmod{n}$.

Suppose that also $d \equiv a \pmod{m}$ and $d \equiv b \mod n$. Then $d \equiv c \pmod{m}$ and $d \equiv c \pmod{n}$, that is $m \mid d - c$ and $n \mid d - c$. Since $\gcd(m, n) = 1$ we have $mn \mid d - c$, that is $d \equiv c \pmod{mn}$. □

To put it differently, the map

$$\mathbb{Z}/mn \to \mathbb{Z}/m \times \mathbb{Z}/n, \quad [c]_{mn} \mapsto ([c]_m, [c]_n)$$

is bijective if $\gcd(m, n) = 1$. The set $\mathbb{Z}/m \times \mathbb{Z}/n$ has mn elements, as does the set $\mathbb{Z}/mn$. The surjectivity, the first part of the theorem, implies in fact the injectivity, the second part of the theorem. Alternatively, one only proves the injectivity, which is even more simple. The theorem implies that computation in $\mathbb{Z}/mn$ can be done in $\mathbb{Z}/m \times \mathbb{Z}/n$. An element $[c]_{mn} \in \mathbb{Z}/mn$ then corresponds to the element $([c]_m, [c]_n) \in \mathbb{Z}/m \times \mathbb{Z}/n$. Addition, multiplication and exponentiation in $\mathbb{Z}/m \times \mathbb{Z}/n$ are done component wise. Furthermore, $[c]_{mn}$ is invertible if and

only if both $[c]_m$ and $[c]_n$ are invertible. In other words, restriction of the map to the invertible elements yields a bijection

$$(\mathbb{Z}/mn)^* \to (\mathbb{Z}/m)^* \times (\mathbb{Z}/n)^*, \quad [c]_{mn} \mapsto ([c]_m, [c]_n).$$

13.26 Example. Figure 13.2 consists of pictures of the permutations

- $\sigma_2 \colon (\mathbb{Z}/35)^* \to (\mathbb{Z}/35)^*,\ \overline{a} \mapsto \overline{2a}$,
- $(\mathbb{Z}/5)^* \times (\mathbb{Z}/7)^* \to (\mathbb{Z}/5)^* \times (\mathbb{Z}/7)^*,\ (\overline{a}, \overline{b}) \mapsto (\overline{2a}, \overline{2b})$,
- $\sigma_2 \colon (\mathbb{Z}/5)^* \to (\mathbb{Z}/5)^*,\ \overline{a} \mapsto \overline{2a}$,
- $\sigma_2 \colon (\mathbb{Z}/7)^* \to (\mathbb{Z}/7)^*,\ \overline{a} \mapsto \overline{2a}$.

The first two permutations correspond via the Chinese Remainder Theorem. The last two permutations determine the permutation of $(\mathbb{Z}/5)^* \times (\mathbb{Z}/7)^*$. It is easily seen that the order of 2 modulo 5 equals 4 and that the order modulo 7 equals 3. So modulo 35 that order equals $\operatorname{lcm}(4, 3) = 12$. For this there is no need to compute the powers of $\overline{2}$ in $(\mathbb{Z}/35)^*$.

From the Chinese Remainder Theorem it follows that the number of invertible elements in $\mathbb{Z}/mn$ is equal to the number of invertible elements in $\mathbb{Z}/m \times \mathbb{Z}/n$ under component wise multiplication. So in particular we have:

13.27 Corollary. *Let $m, n \in \mathbb{N}^+$ satisfy $\gcd(m, n) = 1$. Then $\varphi(mn) = \varphi(m)\varphi(n)$.* □

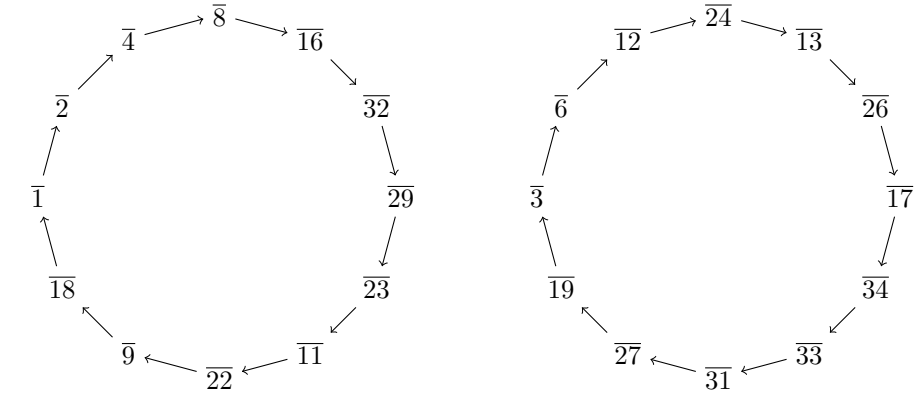
A consequence is the formula for the totient which we already derived in chapter 10:

13.28 Corollary. *Let $m \in \mathbb{N}^+$. Then $\varphi(m) = m \prod_{p|m}(1 - \frac{1}{p})$.*

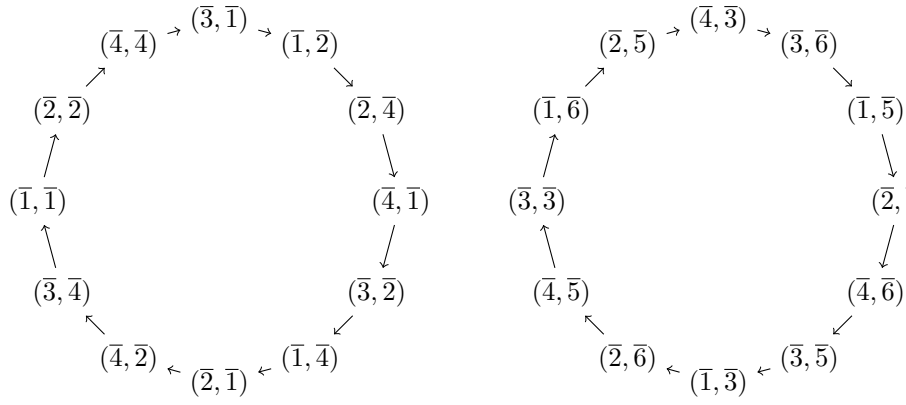
Proof.

$$\begin{aligned}\varphi(m) &= \varphi\Big(\prod_{p|m} p^{v_p(m)}\Big) = \prod_{p|m} \varphi\big(p^{v_p(m)}\big) = \prod_{p|m} \big(p^{v_p(m)} - p^{v_p(m)-1}\big) \\ &= \prod_{p|m} p^{v_p(m)} \cdot \prod_{p|m} \Big(1 - \frac{1}{p}\Big) = m \prod_{p|m} \Big(1 - \frac{1}{p}\Big).\end{aligned}$$ □

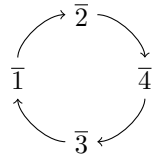
13.29 Example. Again we compute $\overline{2}^{1000}$ in $\mathbb{Z}/45$, now using the Chinese Remainder Theorem. To that end we first compute $\overline{2}^{1000}$ in $\mathbb{Z}/5$ and also in $\mathbb{Z}/9$. In $\mathbb{Z}/5$ we have $\overline{2}^{1000} = \overline{1}$ since $\varphi(5) \mid 1000$. Furthermore $\varphi(9) = 6$ and $1000 = 166 \cdot 6 + 4$, so in $\mathbb{Z}/9$: $\overline{2}^{1000} = \overline{2}^4 = \overline{16} = \overline{7}$. So in $\mathbb{Z}/45$: $\overline{2}^{1000} = \overline{16}$, since $16 \equiv 1 \pmod 5$ and $16 \equiv 7 \pmod 9$. This way it is clear that in $\mathbb{Z}/45$ we even have $\overline{a}^{12} = \overline{1}$ for all $\overline{a} \in (\mathbb{Z}/45)^*$, because $[a]_5^{12} = [1]_5$ and $[a]_9^{12} = [1]_9$ (note that $\varphi(5), \varphi(9) \mid 12$).



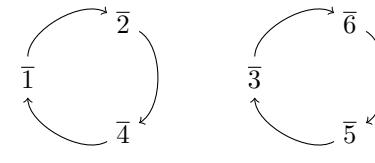
The permutation $\overline{a} \mapsto \overline{2a}$ of $(\mathbb{Z}/35)^*$



The permutation $(\overline{a},\overline{b}) \mapsto (\overline{2a},\overline{2b})$ of $(\mathbb{Z}/5)^* \times (\mathbb{Z}/7)^*$



The permutation $\overline{a} \mapsto \overline{2a}$ of $(\mathbb{Z}/5)^*$



The permutation $\overline{a} \mapsto \overline{2a}$ of $(\mathbb{Z}/7)^*$

Figure 13.2: The permutation σ_2 of $(\mathbb{Z}/35)^*$ and the Chinese Remainder Theorem

In order to determine the order of a modulo mn, the least $k \in \mathbb{N}$ such that $([a]_m^k, [a]_n^k) = ([1]_m, [1]_n)$ can be determined. It is the least common divisor of $\mathrm{o}_m(a)$ and $\mathrm{o}_n(a)$.

13.30 Example. In example 13.12 we saw that $\overline{7}^{10} = \overline{1}$ in $\mathbb{Z}/33$. Instead of exponentiation in $(\mathbb{Z}/33)^*$ we can do exponentiation in $(\mathbb{Z}/3)^* \times (\mathbb{Z}/11)^*$. The order of an $\overline{a}$ in $(\mathbb{Z}/3)^*$ is 1 or 2. In $(\mathbb{Z}/11)^*$ the order of an element is a divisor of 10. The order of an integer modulo 33 is therefore a divisor of $\mathrm{lcm}(2, 10) = 10$. The order of 7 modulo 3 is 1 and modulo 11 it is 10.

Isomorphisms

13.31 Definition. If A and B are algebraic structures like groups or rings, then a bijection $f\colon A \to B$ is called an *isomorphism* if it preserves the operations. If A and B are groups than f is called an *isomorphism of groups* or a *group isomorphism* if f preserves the group operation (no matter how it is denoted). If A and B are rings, then f is an *isomorphism of rings* or a *ring isomorphism* if $f(a_1 + a_2) = f(a_1) + f(a_2)$ and $f(a_1 a_2) = f(a_1) f(a_2)$ for all $a_1, a_2 \in A$ and moreover $f(1) = 1$. (We see 1 as an operation $A^0 \to A$).

Groups A and B are called *isomorphic* if there is a group isomorphism from A to B. Similarly, rings A and B are called *isomorphic* if there is a ring isomorphism from A to B. Notation: $A \cong B$.

If we do not require the map f to be a bijection, but still do require it to preserve the operations, then such an f is called a *homomorphism*. In particular we thus have *homomorphisms of groups* and *homomorphisms of rings*.

13.32 Examples.

1. By the Chinese Remainder Theorem the map

$$\mathbb{Z}/mn \to \mathbb{Z}/m \times \mathbb{Z}/n, \quad [c]_{mn} \mapsto ([c]_m, [c]_n)$$

   is bijective if $\gcd(m, n) = 1$ and, as we already remarked, it preserves the operations. It is an isomorphism of rings, $\mathbb{Z}/m \times \mathbb{Z}/n$ being the ring with component wise operations. Therefore, these rings are isomorphic.

2. Restriction of the isomorphism of rings to their invertible elements yields an isomorphism of groups:

$$(\mathbb{Z}/mn)^* \to (\mathbb{Z}/m)^* \times (\mathbb{Z}/n)^*, \quad [c]_{mn} \mapsto ([c]_m, [c]_n).$$

   So these groups are isomorphic.

13.7 Maximal Orders Modulo m

We determine the order of a power $\overline{a}^k$ of an element $\overline{a} \in (\mathbb{Z}/m)^*$, the order of $\overline{a}$ being given.

13.33 Proposition. *Let $m \in \mathbb{N}^+$ and $a \in \mathbb{Z}$ with $\gcd(a,m) = 1$. Let $\mathrm{o}_m(a) = n$. Then for $k \in \mathbb{Z}$:*

$$\mathrm{o}_m(a^k) = \frac{n}{\gcd(k,n)}.$$

Proof. We determine the integers l satisfying $(\overline{a}^k)^l = \overline{1}$. Equivalent are:

$(\overline{a}^k)^l = \overline{1}$.

$\overline{a}^{kl} = \overline{1}$.

$n \mid kl$.

$\dfrac{n}{\gcd(k,n)} \;\Big|\; \dfrac{k}{\gcd(k,n)} l$.

$\dfrac{n}{\gcd(k,n)} \;\Big|\; l \qquad \Big(\text{since } \gcd\Big(\dfrac{n}{\gcd(k,n)}, \dfrac{k}{\gcd(k,n)}\Big) = 1\Big)$.

From this it follows that $\mathrm{o}_m(a^k) = \dfrac{n}{\gcd(k,n)}$. □

13.34 Corollary. *Let $m \in \mathbb{N}^+$ and $a \in \mathbb{Z}$ with $\gcd(a,m) = 1$. Let $\mathrm{o}_m(a) = n$ and $d \in \mathbb{N}$ with $d \mid n$. Then*

$$\mathrm{o}_m(a^{\frac{n}{d}}) = d.$$

Proof. Apply proposition 13.33: $\mathrm{o}_m(a^{\frac{n}{d}}) = \frac{n}{\gcd(n,n/d)} = \frac{n}{n/d} = d$. □

If two numbers occur as orders modulo m, then so does their least common multiple. To see this, we first treat a special case.

13.35 Lemma. *Let $m \in \mathbb{N}^+$ and $a_1, a_2 \in \mathbb{Z}$ with $\gcd(a_1,m) = \gcd(a_2,m) = 1$. Let $\mathrm{o}_m(a_1) = n_1$, $\mathrm{o}_m(a_2) = n_2$ and $\gcd(n_1,n_2) = 1$. Then $\mathrm{o}_m(a_1a_2) = n_1n_2$.*

Proof. We determine the numbers $l \in \mathbb{Z}$ which have the property that the l-th power of $\overline{a_1a_2}$ in $(\mathbb{Z}/m)^*$ equals $\overline{1}$. Equivalent are:

$(\overline{a_1a_2})^l = \overline{1}$.

$\overline{a_1}^l = \overline{a_2}^{-l}$.

$\overline{a_1}^l = \overline{a_2}^{-l} = \overline{1}$ (since $\mathrm{o}_m(a_1^l) \mid n_1$, $\mathrm{o}_m(a_2^l) \mid n_2$ and $\gcd(n_1,n_2) = 1$).

$\overline{a_1}^l = \overline{1}$ and $\overline{a_2}^l = \overline{1}$.

$n_1 \mid l$ and $n_2 \mid l$.

$n_1n_2 \mid l$.

So $o_m(a_1a_2) = n_1n_2$. □

13.36 Proposition. *Let $m \in \mathbb{N}^+$ and $a_1, a_2 \in \mathbb{Z}$ with $\gcd(a_1, m) = \gcd(a_2, m) = 1$. Let $o_m(a_1) = n_1$ and $o_m(a_2) = n_2$. Then there exists an integer b such that $\gcd(b, m) = 1$ and $o_m(b) = \operatorname{lcm}(n_1, n_2)$.*

Proof. The number $\operatorname{lcm}(n_1, n_2)$ is a product of prime powers each being a divisor of n_1 or n_2. Let p^k be one of these prime powers, say $p^k \mid n_1$. By Corollary 13.34 there exists an integer a_p such that $o_m(a_p) = p^k$. By lemma 13.35 for the product b of the integers a_p we have $o_m(b) = \operatorname{lcm}(n_1, n_2)$. □

Using proposition 13.36 we derive that all orders occurring modulo m divide the maximal order modulo m.

13.37 Proposition. *Let $m \in \mathbb{N}^+$ and let N be the maximal order modulo m. Then $\overline{a}^N = \overline{1}$ for all $\overline{a} \in (\mathbb{Z}/m)^*$.*

Proof. Choose $a_0 \in \mathbb{Z}$ with $\gcd(a_0, m) = 1$ and $o_m(a_0) = N$. Let $a \in \mathbb{Z}$ with $\overline{a} \in (\mathbb{Z}/m)^*$. Then $o_m(a) \le N$, because N is the maximal order. By proposition 13.36 there is an element $\overline{b}$ with $o_m(b) = \operatorname{lcm}(o_m(a), N)$. Then also $o_m(b) \le N$, that is $\operatorname{lcm}(o_m(a), N) \le N$. From this it follows that $o_m(a) \mid N$. So $\overline{a}^N = \overline{1}$. □

Primitive roots modulo p

From proposition 13.37 it will follow that to each prime number p there is a number g of order $p-1$ modulo p. For such a g there are $p-1$ different powers of $\overline{g} \in \mathbb{F}_p$, that is

$$\mathbb{F}_p^* = \{\overline{1}, \overline{g}, \overline{g}^2, \ldots, \overline{g}^{p-2}\}.$$

13.38 Definition. Let $m \in \mathbb{N}^+$. An $a \in \mathbb{Z}$ with $\gcd(a, m) = 1$ is called a *primitive root modulo m* if $o_m(a) = \varphi(m)$.

> Not for all m does there exist a primitive root modulo m. There is no primitive root modulo 8: $o_8(1) = 1$ and $o_8(3) = o_8(5) = o_8(7) = 2$.

13.39 Theorem. *Let p be a prime number. Then there exists a primitive root modulo p.*

Proof. Let N be the maximal order modulo p. Then $N \mid p-1$. We have to prove that $N = p-1$. From proposition 13.37 follows that every $\overline{a} \in \mathbb{F}_p^*$ satisfies $\overline{a}^N = \overline{1}$. So each of the $p-1$ elements $\overline{a} \in \mathbb{F}_p^*$ is a solution of the equation $x^N - \overline{1}$. Because $\mathbb{F}_p$ is a field this equation has at most N solutions. So $p-1 \le N$. Since also $N \mid p-1$, we have $N = p-1$. □

If g is a primitive root modulo a prime number p, then the map $\mathbb{Z}/(p-1) \to \mathbb{F}_p^*$, $\overline{k} \mapsto \overline{g}^k$ is bijective. It is an isomorphism of groups: $\overline{k} + \overline{l} \mapsto \overline{g}^k \overline{g}^l$. The group operation in $\mathbb{Z}/(p-1)$ is the addition and in $\mathbb{F}_p^*$ it is the multiplication. This isomorphism is an exponential map. Its inverse is a kind of logarithm.

13.40 Examples. Finding a primitive root modulo a prime number p is not always easy. For small p you can simply try. For $p = 17$ for example you try 2 (doing arithmetic modulo 17):

$$2^2 = 4, \quad 2^3 = 8, \quad 2^4 = 16 \equiv -1$$

and clearly $2^8 \equiv 1$. So $\mathrm{o}_{17}(2) = 8$. Next you try 3. Since the order divides 16, it is convenient to square repeatedly:

$$3^2 = 9 \equiv -8, \quad 3^4 \equiv 64 \equiv -4, \quad 3^8 \equiv 16 \equiv -1,$$

so $\mathrm{o}_{17}(3) = 16$, that is 3 is a primitive root modulo 17.

When looking for a primitive root modulo a prime number p it helps when the prime factorization of $p-1$ is known, see Corollary 13.42 and example 13.43.

13.41 Lemma. *Let $n \in \mathbb{N}^+$ and $a \in \mathbb{Z}$ be such that $\gcd(a, n) = 1$. Suppose $k \in \mathbb{N}^+$ satisfies $a^k \equiv 1 \pmod{n}$. Then*

$$\mathrm{o}_n(a) = k \iff a^{\frac{k}{q}} \not\equiv 1 \pmod{n} \text{ for all prime divisors } q \text{ of } k.$$

Proof. From $a^k \equiv 1 \pmod{n}$ follows that $\mathrm{o}_n(a) \mid k$. Equivalent are:

$\mathrm{o}_n(a) \neq k$.

There is a prime divisor q of $\dfrac{k}{\mathrm{o}_n(a)}$.

There is a prime divisor q of k with $\mathrm{o}_n(a) \mid \frac{k}{q}$.

There is a prime divisor q of k with $a^{\frac{k}{q}} \equiv 1 \pmod{n}$. □

13.42 Corollary. *Let p be a prime number and $g \in \mathbb{Z}$ such that $p \nmid g$. Then g is a primitive root modulo p if and only if*

$$g^{\frac{p-1}{q}} \not\equiv 1 \pmod{p} \quad \text{for all prime divisors } q \text{ of } p-1.$$

Proof. From $p \nmid g$ follows that $\gcd(g, p) = 1$. By Fermat's Little Theorem we have $g^{p-1} \equiv 1 \pmod{p}$. Then apply lemma 13.41 with $a = g$, $n = p$ and $k = p - 1$. □

Fast exponentiation while doing arithmetic modulo p makes this proposition worthwhile in practice.

13.43 Example. We look for a primitive root modulo the prime number 5441. The factorization of 5440 is easy because of the many factors 2. We have $5440 = 2^6 \cdot 5 \cdot 17$. First we try 2: from $2^{\frac{5440}{2}} = 2^{2720} \equiv 1 \pmod{5441}$ follows that 2 is not a primitive root. We try 3. We get (with the method of section 13.3):

$$3^{\frac{5440}{2}} = 3^{2720} \equiv 5440, \quad 3^{\frac{5440}{5}} = 3^{1088} \equiv 1685, \quad 3^{\frac{5440}{17}} = 3^{320} \equiv 2670,$$

So 3 is a primitive root modulo 5441.

Roots of unity

Solutions of equations of type $x^m - 1 = 0$ are called roots of unity. In the field $\mathbb{Q}$ there are only two of them, 1 and -1, but for fields $\mathbb{F}_p$ it is different. The roots of unity terminology is introduced here mainly for later use.

13.44 Definition. Let K be a field and $m \in \mathbb{N}^+$. A $\zeta \in K$ is called an *m-th root of unity* if $\zeta^m = 1$ and a *primitive* m-th root of unity if moreover $\zeta^k \neq 1$ for all $k \in \mathbb{N}^+$ with $k < m$.

> Instead of $\zeta^m = 1$ you might write $\zeta = \sqrt[m]{1}$, but then you have to realize that in general ζ is not determined this way. It does explain the terminology: it is a root of 1, the unity element of the field.

13.45 Examples.

a) The field $\mathbb{Q}$ has two roots of unity: 1 and -1. These are 2nd roots of unity and -1 is the only primitive 2nd root of unity.
b) The field $\mathbb{F}_2$ has only one root of unity and that is $\overline{1}$.
c) The field $\mathbb{F}_5$ has four roots of unity. The elements 2 and 3 are primitive 4th roots of unity.

Let p be a prime number. For every $\overline{a} \in \mathbb{F}_p^*$ we have $\overline{a}^{p-1} = \overline{1}$. So the elements of $\mathbb{F}_p^*$ are $(p-1)$-st roots of unity. If the order of a modulo p equals k, that is $\mathrm{o}_p(a) = k$, then $\overline{a} \in \mathbb{F}_p$ is a primitive k-th root of unity. Theorem 13.39 tells us that in $\mathbb{F}_p$ there is a primitive $(p-1)$-st root of unity.

Roots of unity have a role in solving equations of type $x^m - a = 0$ with $a \neq 0$. If such an equation has a solution $x = b$ and if the field has a primitive m-th root of unity ζ, then the equation has m solutions: the elements $b, \zeta b, \zeta^2 b, \ldots, \zeta^{m-1}$ are m different zeros of the polynomial $x^m - a$. There are no more zeros because the degree is m. Therefore,

$$x^m - a = (x - b)(x - \zeta b)(x - \zeta^2 b) \cdots (x - \zeta^{m-1} b).$$

See also example 13.18. In chapter 18 we will determine the roots of unity in the field $\mathbb{Q}_p$ of the p-adic numbers and in chapter 19 those in the field $\mathbb{C}$ of the complex numbers. In $\mathbb{C}$ there are plenty roots of unity: for each $m \in \mathbb{N}^+$ the number of m-th roots of unity is m.

Exercises

1. Solve in $\mathbb{F}_{167}$: $\overline{41} \cdot x = \overline{13}$.
2. Prove using modular arithmetic: $13 \mid 8 \cdot 51^n + 5 \cdot 64^n$ for all natural numbers n.
3. Let p be a prime number.
   (i) Determine all $x \in \mathbb{F}_p$ satisfying $x^2 = \overline{1}$.
   (ii) Prove Wilson's theorem: $(p-1)! \equiv -1 \pmod{p}$.
4. Determine the least natural number x satisfying:
   $$x \equiv n - 1 \pmod{n}$$
   for all natural numbers n with $2 \leq n \leq 10$.
5. Compute $\#((\mathbb{Z}/270)^*)$.
6. Solve in $\mathbb{Z}/16$: $x^2 = \overline{1}$.
7. In the decimal notation of a natural number it is easy to see whether it is divisible by 2 or 5. Divisibility by 3, 9 or 11 is also easy. Let $n \in \mathbb{N}^+$ and let its decimal notation be $\cdots a_3 a_2 a_1 a_0$. Show that
   (i) $3 \mid n \iff 3 \mid \sum_i a_i$.
   (ii) $9 \mid n \iff 9 \mid \sum_i a_i$.
   (iii) $11 \mid n \iff 11 \mid \sum_i (-1)^i a_i$.
8. Compute $\overline{2}^{1000000}$ in $\mathbb{Z}/55$.
9. Prove that $63 \mid n^7 - n$ for all $n \in \mathbb{Z}$ with $3 \nmid n$.
10. Let $m \in \mathbb{N}^+$ and $a \in \mathbb{Z}$ with $\gcd(a, m) = 1$. An alternative proof of Euler's theorem (theorem 13.22) is as follows . Let N be the product of all $k \in \mathbb{N}_m$ with $\gcd(k, m) = 1$.
    (i) Show that $\overline{N} \in (\mathbb{Z}/m)^*$.
    (ii) Let M be the product of all ak where $\gcd(k, m) = 1$. Show that $M = a^{\varphi(m)} N$.
    (iii) Show that $\overline{N} = \overline{M}$.
    (iv) Finish the proof of Euler's theorem.
11. Let p be an odd prime number. The permutations σ and τ of $\mathbb{F}_p^*$ are defined by $\sigma(x) = -x$ and $\tau(x) = x^{-1}$ (for all $x \in \mathbb{F}_p^*$).
    (i) Show that σ, τ and $\sigma\tau$ are products of disjoint transpositions (2-cycles).
    (ii) Determine $\operatorname{sgn}(\sigma)$, $\operatorname{sgn}(\tau)$ and $\operatorname{sgn}(\sigma\tau)$.
    (iii) Show that there are 0 or 2 elements $y \in \mathbb{F}_p^*$ satisfying $y^2 = -\overline{1}$.
    (iv) Prove that $-\overline{1}$ is a square in $\mathbb{F}_p$ if and only if $p \equiv 1 \pmod{4}$.
12. Let p be an odd prime number and $\overline{a} \in \mathbb{F}_p^*$. The permutations σ and τ of $\mathbb{F}_p^*$ are defined by $\sigma(x) = \overline{a}x$ and $\tau(x) = x^{-1}$ (for all $x \in \mathbb{F}_p^*$)

(i) Show that $\sigma\tau$ is a product of disjoint transpositions.

(ii) Show that there are 0 or 2 elements $y \in \mathbb{F}_p^*$ with $y^2 = \overline{a}$.

(iii) Prove that σ is an even permutation if and only if $\mathrm{o}_p(a) \mid \frac{p-1}{2}$.

(iv) Prove that $\overline{a}$ is a square in $\mathbb{F}_p$ if and only if $\mathrm{o}_p(a) \mid \frac{p-1}{2}$.

13. By the Chinese Remainder Theorem exponentiation in $\mathbb{Z}/m$ corresponds to simultaneous exponentiation in $\mathbb{Z}/p^{v_p(m)}$ for each prime divisor p of m. For exponentiation in a $\mathbb{Z}/p^r$, where p is a prime number and $r \in \mathbb{N}^+$, there are two cases.

(i) Let $a \in \mathbb{Z}$ with $p \nmid a$. Show that $a^{p^{r-1}(p-1)} \equiv 1 \pmod{p^r}$.

(ii) Let $a \in \mathbb{Z}$ with $p \mid a$. Show that there is an $N \in \mathbb{N}$ with $a^N \equiv 0 \pmod{p^r}$. What is the least N having this property?

14. Let p and q be different odd prime numbers. What is the maximal order in $(\mathbb{Z}/pq)^*$? (Hint: use the Chinese Remainder Theorem.)

15. Let p be an odd prime number.

(i) Show that there is a primitive root modulo $2p$.

(ii) Show that there is a primitive root modulo $4p$.

(iii) Show that there is no primitive root modulo $8p$.

16. (i) Prove that $v_2(5^{2^n} - 1) = n + 2$ for all $n \in \mathbb{N}$.

(ii) Let $n \in \mathbb{N}^+$. Determine the order of 5 modulo 2^n.

17. We do arithmetic modulo 65. Determine $\overline{3}^{65}$ in three ways:

(i) by exponentiation while doing arithmetic modulo 65,

(ii) by using Euler's theorem,

(iii) by using the Chinese Remainder Theorem.

18. Again exercise 11(iv). Now using the existence of a primitive root modulo a prime number. Let p be an odd prime number.

(i) Prove that $\overline{-1}$ is a square in $\mathbb{F}_p$ if and only if there is an $a \in \mathbb{Z}$ with $\mathrm{o}_p(a) = 4$.

(ii) Prove that there is an $a \in \mathbb{Z}$ with $\mathrm{o}_p(a) = 4$ if and only if $p \equiv 1 \pmod{4}$.

19. Let p a be an odd prime number and g a primitive root modulo p.

(i) Show that $p - 1 \mid \mathrm{o}_{p^2}(g)$.

(ii) Show that $\mathrm{o}_{p^2}(g^p) = p - 1$.

(iii) Show that $\mathrm{o}_{p^2}(p + 1) = p$.

(iv) Prove that there is a primitive root modulo p^2.

20. (i) Determine the remainder of $5^{72727272}$ after division by 72.

(ii) Determine the remainder of $2^{72727272}$ after division by 72.

21. Let p and q be different odd prime numbers. Prove that $2^{\frac{(p-1)(q-1)}{2}} \equiv 1 \pmod{pq}$.

22. (i) Show that $2^{5^{36}} \equiv 2 \pmod{109}$.
    (ii) Determine the remainder of $2^{6^{36}}$ after division by 109.

23. Let p and q be different prime numbers. Prove that
$$p^{q-1} + q^{p-1} \equiv 1 \pmod{pq}.$$

24. Let p be a prime number and $k \in \mathbb{N}$ with $k \leq p$. Prove that
$$\binom{p-1}{k} \equiv (-1)^k \pmod{p}.$$

25. For $a, b \in \mathbb{Z}$ we define a transformation $\tau_{a,b} \colon \mathbb{F}_5 \to \mathbb{F}_5$ by $\tau_{a,b}(x) = \overline{a}x + \overline{b}$.
    (i) Prove that $\tau_{a,b}$ is a permutation if and only if $5 \nmid a$. How many permutations of $\mathbb{F}_5$ are there of this type?
    (ii) Let G be the set of permutations $\tau_{a,b}$ of $\mathbb{F}_5$ with $a, b \in \mathbb{Z}$ and $5 \nmid a$. Let $\sigma \in G$. Show that $\sigma^{-1} \in G$.
    (iii) Prove that for all $m \in \mathbb{N}$
$$\tau_{2,1}^m = 1 \iff 4 \mid m.$$

26. Let $a \in \mathbb{N}^+$ and p a prime number. The transformation σ of the set $\underline{a}^p$ is defined by
$$\sigma(a_1, a_2, \ldots, a_p) = (a_2, \ldots, a_p, a_1) \qquad (\text{for } a_1, \ldots, a_p \in \underline{a}).$$
    (i) Show that σ is a permutation with orbits having 1 and p elements.
    (ii) How many elements of $\underline{a}^p$ belong to an orbit of p elements?
    (iii) Prove Fermat's Little Theorem using the previous part of this exercise.

27. We consider the sequence $f_0, f_1, f_2, \ldots$ of Fibonacci numbers (with $f_0 = 0$ and $f_1 = 1$). Given is an $m \in \mathbb{N}^+$.
    (i) Prove that for all $k \in \mathbb{N}$:
$$f_{m+k} \equiv f_{m-1} f_k \pmod{f_m}.$$
    (ii) Prove that for all $l \in \mathbb{N}^+$:
$$f_{lm} \equiv f_{m-1} f_{(l-1)m} \pmod{f_m}.$$
    (iii) Prove that for all $l \in \mathbb{N}$ we have $f_m \mid f_{lm}$.

14 Quadratic Residues

A well-known problem is: which natural numbers are representable as the sum of two squares? In principle it is possible to determine whether a given $n \in \mathbb{N}^+$ is such a sum since there are only finitely many squares $\leq n$. We will show that the existence of a solution depends on the prime factorization of n. This was stated by Fermat, but, which is less known, also the French mathematician **Albert Girard** (1595-1632) did so before Fermat. It was not his habit to publish, but later others, among them Euler, did publish proofs for Fermat's results. A more general problem is about writing numbers in the form $x^2 - ay^2$ with $x, y \in \mathbb{Z}$, where a is a given integer. For $a = -1$ this is the sum of two squares problem. In section 14.1 we will show that this 'representation problem' leads to the question: for which odd prime numbers p is $\overline{a}$ a square in the field $\mathbb{F}_p$? In the sections 14.2 up to 14.7 the theory of squares in fields $\mathbb{F}_p$ is treated. It culminates in the Quadratic Reciprocity Law (section 14.5). In section 14.7 a technique for extracting square roots of squares in $\mathbb{F}_p$ is discussed. In the last section the theory of quadratic residues is applied to the aforementioned representation problems. In the next chapter a completely different application is given: the determination whether a number is prime or composite.

14.1 Representation by Quadratic Forms (1)

14.1 Definition. Let R be a commutative ring. A polynomial in x and y of type $ax^2 + bxy + cy^2$ with $a, b, c \in R$ and $(a, b, c) \neq (0, 0, 0)$ is called a *quadratic form* in x and y over R. Quadratic forms over $\mathbb{Z}$ are also called *integral* quadratic forms and those over $\mathbb{Q}$ *rational* quadratic forms.

> More generally a *form of degree* d is a homogeneous polynomial of degree d. A polynomial in $x_1, \ldots, x_n$ being *homogeneous of degree* d if it is a sum of terms $ax_1^{k_1} \cdots x_n^{k_n}$ with $k_1 + \cdots + k_n = d$ and a an element of the ring.

In this chapter we consider only integral quadratic forms $x^2 - ay^2$ with $a \neq 0$. Gauß developed a beautiful theory for integral quadratic forms in general, but this theory is not in the scope of this book.

14.2 Definition. Let $a \in \mathbb{Z}$, not a square. We say that an $n \in \mathbb{Z}$ is *representable* by the form $x^2 - ay^2$ if there are $x, y \in \mathbb{Z}$ such that $x^2 - ay^2 = n$.

For a given a we will study the following problem:

> Which $n \in \mathbb{Z}$ are representable by the form $x^2 - ay^2$?

Thus for every nonsquare a we have a *representation problem*. For $a = -1$ this is the problem

> Which integers are the sum of two squares?

In this section we will solve this last problem. We will see that the representability of an n as a sum of two squares depends on its prime factorization. A first important step is a remark by Euler:

14.3 Lemma (Euler). *If m and n are representable by the form $x^2 - ay^2$, then so is mn.*

Proof. This follows from the identity

$$(x^2 - ay^2)(u^2 - av^2) = (xu + ayv)^2 - a(yu + xv)^2. \qquad \square$$

The following notion was introduced by Euler.

14.4 Definition. Let $a \in \mathbb{Z}$ be a nonsquare. An odd prime number p is called an *essential prime divisor* of the form $x^2 - ay^2$ if there are $x, y \in \mathbb{Z}$ such that $p \nmid a$, $\gcd(x, y) = 1$ and $p \mid x^2 - ay^2$.

> Note that a common divisor of x and y is a divisor of $x^2 - ay^2$. A divisor of a is a divisor of $0^2 - a \cdot 1^2$ and 2 is a divisor of $1^2 - a \cdot 1^2$ if a is odd.

14.5 Proposition. *Let $a \in \mathbb{Z}$ with $a \neq 0$ and let p be an odd prime number. Then the following are equivalent:*

(i) *p is an essential prime divisor of the form $x^2 - ay^2$,*
(ii) *$\overline{a}$ is a square in $\mathbb{F}_p$.*

Proof.

(i)⇒(ii) There are $x, y \in \mathbb{Z}$ with $\gcd(x, y) = 1$ and $p \mid x^2 - ay^2$. Then in $\mathbb{F}_p$ we have $\overline{x}^2 = \overline{a} \cdot \overline{y}^2$. Moreover $\overline{y} \neq 0$, because otherwise $p \mid \gcd(x, y)$. So $\overline{a} = (\overline{x} \cdot \overline{y}^{-1})^2$.

(ii)⇒(i) There is an $x \in \mathbb{Z}$ with $\overline{a} = \overline{x}^2$ in $\mathbb{F}_p$. From $p \nmid a$ follows $\overline{x} \neq \overline{0}$. Since $a \equiv x^2 \pmod{p}$, there is an integer k such that $a = x^2 + kp$, that is $x^2 - a \cdot 1^2 = -kp$. $\square$

14.6 Example. Though $\overline{-5}$ is a square in $\mathbb{F}_3$, the number 3 is not representable by the form $x^2 + 5y^2$. If a prime number p is representable by a form $x^2 - ay^2$, then p is an essential prime divisor of that form. The converse does not hold.

Axel Thue (Tønsberg 1863 – Oslo 1922)

The Norwegian mathematician Thue contributed significantly to the theory of Diophantine equations. Other important work concerned the theory of 'semigroups' (sets with an associative operation). He was professor in applied mathematics. A quote from him on applied mathematics: "The further removed from usefulness or practical application, the more important."

So the essential prime divisors of the form $x^2 - ay^2$ are the odd prime numbers p for which $\overline{a}$ is a square in $\mathbb{F}_p^*$. At first sight the determination of these essential prime divisors seems to be impossible: there is an infinity of prime numbers. Nevertheless a regularity can be discovered. This regularity is given by the Quadratic Reciprocity Law (theorem 14.26), the most important theorem of this chapter.

Sums of two squares

In the exercises 11 and 18 of chapter 13 it is shown in two ways that for odd prime numbers p the residue class $\overline{-1}$ is a square in $\mathbb{F}_p$ if and only if $p \equiv 1 \pmod 4$. In this chapter we will see some more proofs of this: Corollary 14.20 and example 14.23. For the case $a = -1$ this is the kind of regularity we are looking for. Later we will see that there is such a regularity in general.

So by proposition 14.5 the essential prime divisors of the form $x^2 + y^2$ are the prime numbers p with $p \equiv 1 \pmod 4$. In this case these prime numbers not only are essential prime divisors, they are representable by the form $x^2 + y^2$ themselves. We present a proof using Dirichlet's principle (theorem 5.35). The idea of the proof comes from the Norwegian mathematician Thue.

14.7 Lemma. *Let p be a prime number and $s \in \mathbb{Z}$. Then there are $x, y \in \mathbb{Z}$ with $p \mid x^2 - s^2y^2$, x and y not both 0, and $x^2, y^2 < p$.*

Proof. Let t be the natural number satisfying $(t-1)^2 < p < t^2$. The set $\mathbb{N}_t^2$ has t^2 elements, being more than p. Therefore, the map

$$\mathbb{N}_t^2 \to \mathbb{F}_p, \quad (a, b) \mapsto \overline{a + sb}$$

is not injective. So there exist $(a, b), (c, d) \in \mathbb{N}_t^2$ with $(a, b) \neq (c, d)$ and $a + sb \equiv c + sd \pmod p$, that is $a - c \equiv s(d - b) \pmod p$. Put $x = a - c$ and $y = d - b$.

Then $-(t-1) \leq x, y \leq t-1$ and so $x^2, y^2 \leq (t-1)^2 < p$. From $x \equiv sy \pmod{p}$ follows $p \mid x^2 - s^2y^2$. These x and y are not both 0, since $(a,b) \neq (c,d)$. □

14.8 Theorem. *Let p be a prime number with $p \equiv 1 \pmod{4}$. Then there are $x, y \in \mathbb{Z}$ with $p = x^2 + y^2$.*

Proof. There is an $s \in \mathbb{Z}$ with $s^2 \equiv -1 \pmod{p}$. From lemma 14.7 follows that there are x, y in $\mathbb{Z}$ with $p \mid x^2 + y^2$ and $0 < x^2 + y^2 < p + p = 2p$. From this it follows that $x^2 + y^2 = p$. □

Now it is easy to describe the representability of an $n \in \mathbb{N}^+$ by the form $x^2 + y^2$ in terms of the prime factorization of n.

14.9 Theorem. *Let $n \in \mathbb{N}^+$. Then n is representable by the form $x^2 + y^2$ if and only if $v_p(n)$ is even for all prime numbers p with $p \equiv 3 \pmod{4}$.*

Proof. If $v_p(n)$ is even for all prime numbers p with $p \equiv 3 \pmod{4}$, then n is a product of factors of the form

a) prime number p with $p \equiv 1 \pmod{4}$,
b) p^2 with p a prime number $\equiv 3 \pmod{4}$,
c) 2.

Each of these factors is representable by the form x^2+y^2 and so is n by lemma 14.3.

Suppose $n = x^2 + y^2$ for certain $x, y \in \mathbb{Z}$. Let $p \mid n$ with $p \equiv 3 \pmod{4}$. Then to prove that $v_p(n)$ is even. Let $d = \gcd(x, y)$. Then $x = dx_0$ and $y = dy_0$ with $x_0, y_0 \in \mathbb{Z}$ and we have $\gcd(x_0, y_0) = 1$. Then $n = d^2(x_0^2 + y_0^2)$. Since p is not an essential prime divisor and $\gcd(x_0, y_0) = 1$, we have $p \nmid x_0^2 + y_0^2$. So $v_p(n) = v_p(d^2) = 2v_p(d)$. □

14.2 Squares in $\mathbb{F}_p$

In this section p is an odd prime number. The field $\mathbb{F}_p$ consists of the residue classes modulo p:

$$\mathbb{F}_p = \{\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{p-1}\}.$$

Only the element $\overline{0}$ has no inverse:

$$\mathbb{F}_p^* = \{\overline{1}, \overline{2}, \ldots, \overline{p-1}\}.$$

14.10 Definition. Let $a \in \mathbb{Z}$. Then $\overline{a} \in \mathbb{F}_p$ is called a *square* in $\mathbb{F}_p$ if there is a $b \in \mathbb{Z}$ with $\overline{b}^2 = \overline{a}$. We then also say that a is a *square modulo p* or that a is a *quadratic residue modulo p*. Integers which are not squares modulo p are called *nonsquares* modulo p.

14.11 Example. In $\mathbb{F}_{13}^*$ we have:

$$\begin{array}{lcccccccccccc} \overline{a}: & \overline{1} & \overline{2} & \overline{3} & \overline{4} & \overline{5} & \overline{6} & \overline{7} & \overline{8} & \overline{9} & \overline{10} & \overline{11} & \overline{12} \\ \overline{a}^2: & \overline{1} & \overline{4} & \overline{9} & \overline{3} & \overline{12} & \overline{10} & \overline{10} & \overline{12} & \overline{3} & \overline{9} & \overline{4} & \overline{1} \end{array}$$

So the squares in $\mathbb{F}_{13}^*$ are: $\overline{1}, \overline{3}, \overline{4}, \overline{9}, \overline{10}$ and $\overline{12}$. There are 6 of them, that is half of $\#(\mathbb{F}_{13}^*) = 12$.

14.12 Proposition. *If $\overline{a} \in \mathbb{F}_p^*$ is a square, then it is the square of exactly two elements.*

Proof. Let $\overline{a} = \overline{b}^2$. If for $x \in \mathbb{F}_p^*$ we have $x^2 = \overline{a}$, then

$$\overline{0} = x^2 - \overline{a} = x^2 - \overline{b}^2 = (x - \overline{b})(x + \overline{b}),$$

so $x - \overline{b} = \overline{0}$ or $x + \overline{b} = \overline{0}$, that is $x = \overline{b}$ or $x = -\overline{b}$. Since $\overline{b} \neq \overline{0}$ there are two solutions. Note that we used that p is odd: $\overline{b} \neq -\overline{b}$. □

14.13 Proposition. *In $\mathbb{F}_p^*$ there are exactly $\frac{p-1}{2}$ squares.*

Proof. Consider the transformation

$$\mathbb{F}_p^* \to \mathbb{F}_p^*, \quad x \mapsto x^2.$$

There are exactly 2 elements in the inverse image of each of the image elements. So the number of image elements is $\frac{p-1}{2}$. □

So in $\mathbb{F}_p^*$ there are as many squares as there are nonsquares.

By Fermat's Little Theorem $\overline{a}^{p-1} = \overline{1}$ for every $\overline{a} \in \mathbb{F}_p^*$. The power $\overline{a}^{\frac{p-1}{2}}$ determines whether $\overline{a}$ is a square or not:

14.14 Theorem (Euler's Criterion). *For $\overline{a} \in \mathbb{F}_p^*$ we have:*

$$\overline{a} \text{ is a square in } \mathbb{F}_p \iff \overline{a}^{\frac{p-1}{2}} = \overline{1}.$$

Proof. Since $\overline{a} \neq \overline{0}$ we have $\overline{a}^{p-1} - \overline{1} = \overline{0}$. So

$$\left(\overline{a}^{\frac{p-1}{2}} - \overline{1}\right)\left(\overline{a}^{\frac{p-1}{2}} + \overline{1}\right) = \overline{0}.$$

All $\frac{p-1}{2}$ squares are zeros of the polynomial $x^{\frac{p-1}{2}} - \overline{1}$: if $\overline{a} = \overline{b}^2$, then

$$\overline{a}^{\frac{p-1}{2}} - \overline{1} = \overline{b}^{2\frac{p-1}{2}} - \overline{1} = \overline{b}^{p-1} - \overline{1} = \overline{0}.$$

Because the polynomial $x^{\frac{p-1}{2}} - \overline{1}$ is of degree $\frac{p-1}{2}$, it has no more zeros. So a nonsquare is not a zero of this polynomial, but it is a zero of $x^{\frac{p-1}{2}} + \overline{1}$. □

Adrien-Marie Legendre (Paris 1752 – Paris 1832)

The quadratic reciprocity law (see theorem 14.26) was treated by Legendre extensively, however the proof he gave was not complete. As a matter of fact the quadratic reciprocity law was described earlier by Euler (and he too did not prove it, though he came close to a proof). In later work on number theory (*Théorie des nombres*) Legendre gave the proof found by Gauß.

His *Eléments de géométrie* replaced the Elements of Euclid and was since then the basis for many textbooks on geometry. His work on 'elliptic integrals' was of importance for mathematical physics. He showed in a relatively simple way that the number π is not rational.

The only known portrait of Legendre is this caricature. Often erroneously a portrait of the politician and contemporary Louis Legendre is shown.

A direct consequence is:

14.15 Corollary. *For $\overline{a}, \overline{b} \in \mathbb{F}_p^*$ we have: $\overline{ab}$ is a square if and only if $\overline{a}$ and $\overline{b}$ both are squares or both are nonsquares.*

Proof. This follows from $\overline{ab}^{\frac{p-1}{2}} = \overline{a}^{\frac{p-1}{2}} \overline{b}^{\frac{p-1}{2}}$. □

14.3 The Legendre Symbol

The Legendre symbol $\left(\frac{a}{p}\right)$ indicates whether a is a square modulo p.

14.16 Definition. Let p an odd prime number and $a \in \mathbb{Z}$. We define:

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } p \mid a, \\ 1 & \text{if } p \nmid a \text{ and } \overline{a} \text{ is a square in } \mathbb{F}_p, \\ -1 & \text{if } p \nmid a \text{ and } \overline{a} \text{ is not a square in } \mathbb{F}_p. \end{cases}$$

$\left(\frac{a}{p}\right)$ is called a *Legendre symbol*.

By definition the Legendre symbol $\left(\frac{a}{p}\right)$ depends only on the residue class of a modulo p:

14.17 Proposition. *Let p be an odd prime number and $a, b \in \mathbb{Z}$ with $a \equiv b \pmod{p}$. Then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.* □

We will formulate the results on squares in $\mathbb{F}_p$ in terms of Legendre symbols. First a reformulation of Euler's Criterion (theorem 14.14):

14.18 Theorem (Euler's Criterion). *Let $a \in \mathbb{Z}$ and p an odd prime number. Then*

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}. \qquad \square$$

The following proposition is a reformulation of Corollary 14.15:

14.19 Proposition. *Let p be an odd prime number. Then for all $a, b \in \mathbb{Z}$:*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \qquad \square$$

From Euler's Criterion follows:

14.20 Corollary. *Let p be an odd prime number. Then*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}.$$

Proof. This follows from: $\left(\frac{-1}{p}\right) \equiv (-1)^{\frac{p-1}{2}} \pmod{p}$. $\square$

This is equivalent to:

$$-1 \text{ is a square modulo } p \iff p \equiv 1 \pmod{4}.$$

Example 14.11 is the case $p = 13$. Indeed, -1 is a quadratic residue: $-1 \equiv 12 \equiv 5^2 \pmod{13}$.

14.21 Example. We use Euler's Criterion to see whether 3 is a quadratic residue modulo 19. We compute $3^{\frac{19-1}{2}}$:

$$3^{\frac{19-1}{2}} = 3^9 = (3^3)^3 = 27^3 \equiv 8^3 \equiv 512 \equiv 18 \equiv -1 \pmod{19}.$$

So 3 is not a quadratic residue modulo 19.

14.4 Gauß's Criterion

Gauß was the first to prove the quadratic reciprocity law. According to his diary this was on April 8th in 1796. Gauß was very proud of this. He gave it the name *Theorema Aureum*, the golden theorem. Eventually he had six proofs. Nowadays many more proofs are known. One of Gauß's proofs, the third, is based on the

Carl Friedrich Gauß (Brunswick 1777 – Göttingen 1855)

Gauß's influence was enormous, in mathematics, in physics and in astronomy. Already at an early age he showed a regular 17-gon to be constructible with straight edge and compass. Such construction problems dated from the Greek antiquity. With the rise of algebra such problems became feasible. Gauß made many contributions to a variety of subjects: number theory, analysis, differential geometry, geodesics, astronomy, magnetism, optics.

so-called *Gauß's Criterion*. For this we group the elements of $\mathbb{F}_p^*$ in pairs $\{\overline{s}, -\overline{s}\}$. Thus we have a partition of $\mathbb{F}_p^*$:

$$\Phi = \{\, \{\overline{s}, -\overline{s}\} \mid \overline{s} \in \mathbb{F}_p^* \,\}.$$

It is the partition of $\mathbb{F}_p^*$ determined by the map $\mathbb{F}_p^* \to \mathbb{F}_p^*$, $x \mapsto x^2$. Let S be a system of representatives of this partition, for example

$$S = \{\, \overline{1}, \overline{2}, \ldots, \overline{\tfrac{p-1}{2}} \,\}.$$

Thus $\mathbb{F}_p^*$ is partitioned into two halves: S and $-S = \{\, -\overline{s} \mid \overline{s} \in S \,\}$. Multiplication by $\overline{a}$ is a permutation of $\mathbb{F}_p^*$. It induces a permutation of the set Φ. Let $y \in \mathbb{F}_p^*$ be the product of all $\overline{s} \in S$:

$$y = \prod_{\overline{s} \in S} \overline{s}.$$

The set $\overline{a}S$ $(= \{\, \overline{a} \cdot \overline{s} \mid \overline{s} \in S \,\})$ is also a system of representatives. We have:

$$\overline{a}^{\frac{p-1}{2}} y = \overline{a}^{\frac{p-1}{2}} \prod_{\overline{s} \in S} \overline{s} = \prod_{\overline{s} \in S} \overline{a} \cdot \overline{s} = (-\overline{1})^N \prod_{\overline{s} \in S} \overline{s} = (-\overline{1})^N y,$$

where N is the number of the $\overline{s} \in S$ with $\overline{a} \cdot \overline{s} \in -S$, that is $N = \#(\overline{a}S \cap -S)$. because $y \neq \overline{0}$ it follows that $\overline{a}^{\frac{p-1}{2}} = (-1)^N$. Thus now we have, using Euler's Criterion:

14.22 Theorem (Gauß's Criterion). *Let p be an odd prime number. Let S be a subset of $\mathbb{F}_p^*$ such that $S \cup -S = \mathbb{F}_p^*$ and $S \cap -S = \emptyset$. Then for all $a \in \mathbb{Z}$ with $p \nmid a$:*

$$\left(\frac{a}{p}\right) = (-1)^{\#(\overline{a}S \cap -S)}. \qquad \square$$

14.23 Example. Again we compute $\left(\frac{-1}{p}\right)$, now using Gauß's Criterion. Under multiplication by $-\overline{1}$ the subset S maps to $-S$. So:

$$\left(\frac{-1}{p}\right) = (-1)^{\#(-S\cap -S)} = (-1)^{\#(-S)} = (-1)^{\frac{p-1}{2}}.$$

We will compute the Legendre symbol $\left(\frac{2}{p}\right)$ using Gauß' Criterion. We take

$$S = \{\,\overline{1}, \overline{2}, \ldots, \overline{\tfrac{p-1}{2}}\,\}.$$

Then

$$-S = \{\,\overline{\tfrac{p+1}{2}}, \overline{\tfrac{p+3}{2}}, \ldots, \overline{p-1}\,\}.$$

The elements of S are represented by the integers a with $0 < a < \frac{p}{2}$ and those of $-S$ by the integers a with $\frac{p}{2} < a < p$. Then the question becomes: for which of the a with $0 < a < \frac{p}{2}$ does $\frac{p}{2} < 2a < p$ hold? So to determine the number of integers a with $\frac{p}{4} < a < \frac{p}{2}$. We distinguish two cases.

a) For $p \equiv 1 \pmod 4$ these are the integers $\frac{p+3}{4}, \ldots, \frac{p-1}{2}$.
   Their number is $\frac{p-1}{2} - \frac{p-1}{4} = \frac{p-1}{4}$.
b) For $p \equiv 3 \pmod 4$ these are the integers $\frac{p+1}{4}, \ldots, \frac{p-1}{2}$.
   Their number is $\frac{p-1}{2} - \frac{p-3}{4} = \frac{p+1}{4}$.

We only need the parity of these numbers, i.e. whether they are even or odd. We have

$$\frac{p^2-1}{8} = \begin{cases} \frac{p-1}{4}\frac{p+1}{2} \equiv \frac{p-1}{4} \pmod 2 & \text{if } p \equiv 1 \pmod 4, \\ \frac{p-1}{2}\frac{p+1}{4} \equiv \frac{p+1}{4} \pmod 2 & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

So we proved:

14.24 Theorem. *Let p be an odd prime number. Then*

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$ □

Put differently: 2 is a quadratic residue modulo $p \iff p \equiv \pm 1 \pmod 8$.

14.25 Example. We already saw in example 14.21 that 3 is not a quadratic residue modulo 19. Now we use Gauß's Criterion. We look at the pairs $\{\overline{a}, -\overline{a}\}$ and their images under multiplication by $\overline{3}$:

$\overline{a}$	$-\overline{a}$	$\overline{3a}$	$-\overline{3a}$
$\overline{1}$	$-\overline{1}$	$\overline{3}$	$-\overline{3}$
$\overline{2}$	$-\overline{2}$	$\overline{6}$	$-\overline{6}$
$\overline{3}$	$-\overline{3}$	$\overline{9}$	$-\overline{9}$
$\overline{4}$	$-\overline{4}$	$-\overline{7}$	$\overline{7}$
$\overline{5}$	$-\overline{5}$	$-\overline{4}$	$\overline{4}$
$\overline{6}$	$-\overline{6}$	$-\overline{1}$	$\overline{1}$
$\overline{7}$	$-\overline{7}$	$\overline{2}$	$-\overline{2}$
$\overline{8}$	$-\overline{8}$	$\overline{5}$	$-\overline{5}$
$\overline{9}$	$-\overline{9}$	$\overline{8}$	$-\overline{8}$

So $\left(\frac{3}{19}\right) = (-1)^3 = -1$.

14.5 The Quadratic Reciprocity Law

The Quadratic Reciprocity Law relates $\left(\frac{p}{q}\right)$ and $\left(\frac{q}{p}\right)$ for two odd prime numbers p and q. This law makes it possible to decide by means of a simple computation whether an integer is a quadratic residue modulo a given prime. It can also be used to answer questions like: modulo which prime numbers is a given a a square? The proof given here is from Frobenius. It is an elementary proof that uses Gauß's Criterion.

14.26 Theorem (The Quadratic Reciprocity Law). *Let p and q be odd prime numbers with $p \neq q$. Then*

$$\left(\frac{p}{q}\right) \cdot \left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}.$$

Put differently:

$$\left(\frac{q}{p}\right) = \begin{cases} -\left(\frac{p}{q}\right) & \text{if } p \equiv q \equiv 3 \pmod 4, \\ \left(\frac{p}{q}\right) & \text{otherwise.} \end{cases}$$

Proof. Take $S = \{\overline{1}, \ldots, \overline{\frac{p-1}{2}}\} \subseteq \mathbb{F}_p^*$. Then:

$$\#(\overline{q}S \cap -S) =$$
$$\#\{\, x \in \mathbb{N} \mid 1 \leq x \leq \tfrac{p-1}{2} \text{ and there is a } y \in \mathbb{Z} \text{ such that } -\tfrac{p}{2} < qx - py < 0 \,\}.$$

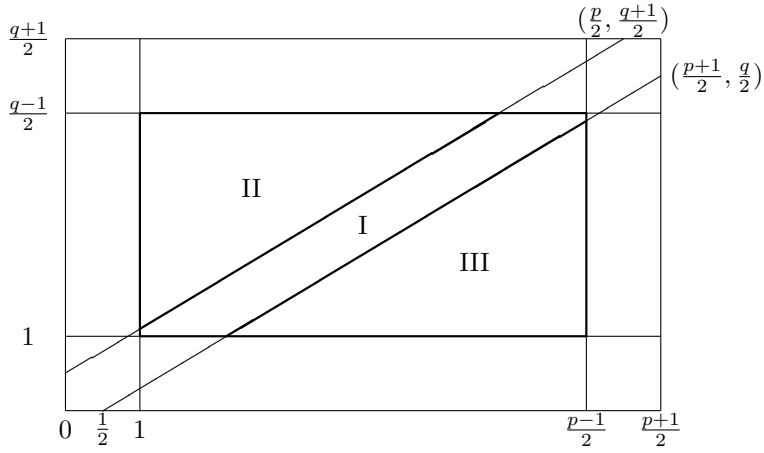


Figure 14.1: To the proof of the quadratic reciprocity law

Note that for each x there is at most one y, and that, if $0 < x < \frac{p}{2}$ and $-\frac{p}{2} < qx - py < 0$, then $qx < py < qx + \frac{p}{2}$ and so $0 < y < \frac{q+1}{2}$, that is $1 \leq y \leq \frac{q-1}{2}$. So:

$$\#(\overline{q}S \cap -S) = \#\{\,(x,y) \in \mathbb{N}^2 \mid 1 \leq x \leq \tfrac{p-1}{2},\ 1 \leq y \leq \tfrac{q-1}{2} \text{ and } -\tfrac{p}{2} < qx - py < 0\,\}.$$

Now for $S' = \{\overline{1}, \ldots, \overline{\frac{q-1}{2}}\} \subseteq \mathbb{F}_q^*$:

$$\begin{aligned}
&\#(\overline{p}S' \cap -S') \\
&\quad = \#\{\,(x,y) \in \mathbb{N}^2 \mid 1 \leq x \leq \tfrac{q-1}{2},\ 1 \leq y \leq \tfrac{p-1}{2} \text{ and } -\tfrac{q}{2} < px - qy < 0\,\} \\
&\quad = \#\{\,(x,y) \in \mathbb{N}^2 \mid 1 \leq x \leq \tfrac{p-1}{2},\ 1 \leq y \leq \tfrac{q-1}{2} \text{ and } 0 < qx - py < \tfrac{q}{2}\,\}.
\end{aligned}$$

Hence: $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^N$ with

$$N = \#\{\,(x,y) \in \mathbb{N}^2 \mid 1 \leq x \leq \tfrac{p-1}{2},\ 1 \leq y \leq \tfrac{q-1}{2} \text{ and } -\tfrac{p}{2} < qx - py < \tfrac{q}{2}\,\}.$$

So we have (see Figure 14.1): $N = \#(\text{lattice points in I})$, and since

$$\#(\text{lattice points in II}) = \#(\text{lattice points in III}),$$

we have

$$N + 2 \cdot \#(\text{lattice points in II}) = \tfrac{p-1}{2} \cdot \tfrac{q-1}{2}. \qquad \square$$

This can also be seen as follows. What matters is the parity of the number of lattice points I. Because of the symmetry of I w.r.t. the middle of the rectangle, this number is odd if and only if this middle is a lattice point. This last is the case if and only if both $\frac{p-1}{2}$ and $\frac{q-1}{2}$ are odd, that is if and only if $p \equiv q \equiv 3 \pmod 4$.

Ferdinand Georg Frobenius (Berlin 1849 – Berlin 1917)

Georg Frobenius was a pupil of Weierstraß. He contributed in an essential way to group theory and number theory. In 1891 he succeeded Kronecker as a professor in Berlin. He had a difficult character, which was not helpful for having good relationships between the institutes of mathematics in Berlin and Göttingen, in those times the most important ones in Germany.

14.27 Example. We determine $\left(\frac{3}{p}\right)$ for prime numbers $p \neq 2, 3$. From the quadratic reciprocity law follows:

$$\left(\frac{3}{p}\right) = (-1)^{\frac{p-1}{2}} \left(\frac{p}{3}\right).$$

The factor $(-1)^{\frac{p-1}{2}}$ depends on p modulo 4 and the factor $\left(\frac{p}{3}\right)$ on p modulo 3. So $\left(\frac{3}{p}\right)$ depends on p modulo 12. There are $\varphi(12) = 4$ cases modulo 12:

$$\left(\frac{3}{p}\right) = \begin{cases} 1 \cdot 1 & \text{if } p \equiv 1 \pmod 4 \text{ and } p \equiv 1 \pmod 3 \\ 1 \cdot -1 & \text{if } p \equiv 1 \pmod 4 \text{ and } p \equiv 2 \pmod 3 \\ -1 \cdot 1 & \text{if } p \equiv 3 \pmod 4 \text{ and } p \equiv 1 \pmod 3 \\ -1 \cdot -1 & \text{if } p \equiv 3 \pmod 4 \text{ and } p \equiv 2 \pmod 3). \end{cases}$$

So:

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 11 \pmod{12}, \\ -1 & \text{if } p \equiv 5, 7 \pmod{12}. \end{cases}$$

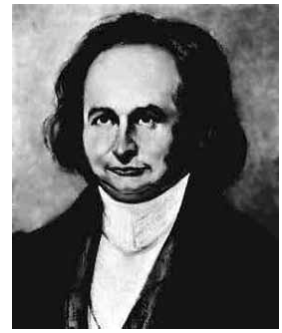
14.28 Example. By repeated application of the quadratic reciprocity law we can determine whether $\overline{47}$ is a square in $\mathbb{F}_{163}$:

$$\begin{aligned} \left(\frac{47}{163}\right) &= -\left(\frac{163}{47}\right) = -\left(\frac{22}{47}\right) = -\left(\frac{2}{47}\right)\left(\frac{11}{47}\right) = -\left(\frac{11}{47}\right) = \left(\frac{47}{11}\right) \\ &= \left(\frac{3}{11}\right) = -\left(\frac{11}{3}\right) = -\left(\frac{2}{3}\right) = 1. \end{aligned}$$

So 47 is a quadratic residue modulo 163. Thus a $b \in \mathbb{Z}$ with $b^2 \equiv 47 \pmod{163}$ is not found. Finding such a square root is another problem. We return to this problem in section 14.7. In fact 79 is a square root of 47 modulo 163.

Carl Jacobi (Potsdam 1804 – Berlin 1851)

Jacobi contributed to number theory already at an early age. Later he also worked on elliptical integrals and differential equations. The 'Jacobian' in the theory of functions of several variables is named after him; he wrote extensively on this, but it had already been introduced by Cauchy.



14.6 The Jacobi Symbol

The Quadratic Reciprocity Law can be used for the computation of Legendre symbols. A complication is that for the application of quadratic reciprocity to $\left(\frac{a}{p}\right)$ the integer a has to be factorized. That can be avoided completely by extending the Legendre symbol to the more general Jacobi symbol for which analogous rules hold.

14.29 Definition. Let a and b be integers with b odd and positive. We define the *Jacobi symbol* $\left(\frac{a}{b}\right)$ by

$$\left(\frac{a}{b}\right) = \prod_p \left(\frac{a}{p}\right)^{v_p(b)}.$$

That is: if $b = p_1 p_2 \cdots p_r$ ($p_1, \ldots, p_r$ being prime numbers), then

$$\left(\frac{a}{b}\right) = \left(\frac{a}{p_1}\right)\left(\frac{a}{p_2}\right)\cdots\left(\frac{a}{p_r}\right).$$

Since b is odd, all prime factors p_i are odd as well. (For $b = 1$ there are 0 prime factors and the product equals 1.)

Simple properties of the Jacobi symbol are:

14.30 Proposition. *Let a, a_1, a_2, b, b_1 and b_2 be integers with b, b_1 and b_2 odd and positive. Then*

(i) $\left(\frac{a}{b}\right) = 0 \iff \gcd(a,b) > 1.$

(ii) *If* $a_1 \equiv a_2 \pmod b$, *then* $\left(\frac{a_1}{b}\right) = \left(\frac{a_2}{b}\right).$

(iii) $\left(\frac{a_1 a_2}{b}\right) = \left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right).$

(iv) $\left(\frac{a}{b_1 b_2}\right) = \left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right)$.

Proof.

(i) From the definition of the Jacobi symbol it follows that $\left(\frac{a}{b}\right) = 0$ if and only if there is a prime divisor p of b such that $\left(\frac{a}{p}\right) = 0$, that is $p \mid a$.

(ii) If $a_1 \equiv a_2 \pmod{b}$, then $a_1 \equiv a_2 \pmod{p}$ for all prime divisors p of b.

(iii)

$$\left(\frac{a_1 a_2}{b}\right) = \prod_p \left(\frac{a_1 a_2}{p}\right)^{v_p(b)} = \prod_p \left(\frac{a_1}{p}\right)^{v_p(b)} \cdot \prod_p \left(\frac{a_2}{p}\right)^{v_p(b)} = \left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right).$$

(iv)

$$\begin{aligned}\left(\frac{a}{b_1 b_2}\right) &= \prod_p \left(\frac{a}{p}\right)^{v_p(b_1 b_2)} = \prod_p \left(\frac{a}{p}\right)^{v_p(b_1)+v_p(b_2)} \\ &= \prod_p \left(\frac{a}{p}\right)^{v_p(b_1)} \cdot \prod_p \left(\frac{a}{p}\right)^{v_p(b_2)} = \left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right).\end{aligned}$$

□

The extension of the rules for the Legendre symbol to the Jacobi symbol rests on the following lemma.

14.31 Lemma. *Let m and n be odd natural numbers. Then*

(i) $\frac{mn-1}{2} \equiv \frac{m-1}{2} + \frac{n-1}{2} \pmod 2$.

(ii) $\frac{(mn)^2-1}{8} \equiv \frac{m^2-1}{8} + \frac{n^2-1}{8} \pmod 2$.

Proof.

(i) Since both $m-1$ and $n-1$ are even, we have $(m-1)(n-1) \equiv 0 \pmod 4$. So $mn-1 \equiv (m-1)+(n-1) \pmod 4$, that is $\frac{mn-1}{2} \equiv \frac{m-1}{2} + \frac{n-1}{2} \pmod 2$.

(ii) Since both m^2-1 and n^2-1 are multiples of 8, we have $(m^2-1)(n^2-1) \equiv 0 \pmod{16}$. So $m^2n^2-1 \equiv (m^2-1)+(n^2-1) \pmod{16}$, that is $\frac{(mn)^2-1}{8} \equiv \frac{m^2-1}{8} + \frac{n^2-1}{8} \pmod 2$. □

These lemmas are about two odd numbers m and n. By induction it follows easily that more generally for odd $m_1, \ldots, m_t$:

$$\begin{aligned}\frac{m_1 \cdots m_t - 1}{2} &\equiv \frac{m_1 - 1}{2} + \cdots + \frac{m_t - 1}{2} \pmod 2 \\ \frac{(m_1 \cdots m_t)^2 - 1}{8} &\equiv \frac{m_1^2 - 1}{8} + \cdots + \frac{m_t^2 - 1}{8} \pmod 2.\end{aligned}$$

We apply this to the prime factorization of an odd number.

14.32 Corollary. *Let a be an odd natural number. Then*

$$\frac{a-1}{2} \equiv \sum_p v_p(a)\frac{p-1}{2} \pmod 2 \quad \text{and} \quad \frac{a^2-1}{8} \equiv \sum_p v_p(a)\frac{p^2-1}{8} \pmod 2.$$

Proof.

$$\frac{a-1}{2} = \frac{\prod_p p^{v_p(a)}-1}{2} \equiv \sum_p v_p(a)\frac{p-1}{2} \pmod 2,$$
$$\frac{a^2-1}{8} = \frac{\prod_p p^{2v_p(a)}-1}{8} \equiv \sum_p v_p(a)\frac{p^2-1}{8} \pmod 2. \qquad \square$$

14.33 Theorem. *Let a and b be odd positive integers with $\gcd(a,b)=1$. Then:*

(i) $\left(\frac{-1}{a}\right) = (-1)^{\frac{a-1}{2}}$.

(ii) $\left(\frac{2}{a}\right) = (-1)^{\frac{a^2-1}{8}}$.

(iii) $\left(\frac{a}{b}\right)\left(\frac{b}{a}\right) = (-1)^{\frac{a-1}{2}\frac{b-1}{2}}$

Proof. We use Corollary 14.32.

(i)

$$\left(\frac{-1}{a}\right) = \prod_p \left(\frac{-1}{p}\right)^{v_p(a)} = \prod_p (-1)^{v_p(a)\frac{p-1}{2}} = (-1)^{\sum_p v_p(a)\frac{p-1}{2}} = (-1)^{\frac{a-1}{2}}.$$

(ii) Analogous to part (i).

(iii)

$$\begin{aligned}\left(\frac{a}{b}\right)\left(\frac{b}{a}\right) &= \prod_q \left(\frac{a}{q}\right)^{v_q(b)} \cdot \prod_p \left(\frac{b}{p}\right)^{v_p(a)} \\ &= \prod_q \prod_p \left(\frac{p}{q}\right)^{v_p(a)v_q(b)} \cdot \prod_p \prod_q \left(\frac{q}{p}\right)^{v_q(b)v_p(a)} \\ &= \prod_p \prod_q \left(\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)\right)^{v_p(a)v_q(b)} = \prod_p \prod_q (-1)^{v_p(a)v_q(b)\frac{p-1}{2}\frac{q-1}{2}} \\ &= (-1)^N\end{aligned}$$

with

$$N = \sum_p \sum_q v_p(a)v_q(b)\frac{p-1}{2}\frac{q-1}{2}$$

$$= \left(\sum_p v_p(a)\frac{p-1}{2}\right)\left(\sum_q v_q(b)\frac{q-1}{2}\right) \equiv \frac{a-1}{2}\frac{b-1}{2} \pmod 2. \quad \square$$

14.34 Example. Is 54321 a quadratic residue modulo the prime number 73673?

$$\begin{aligned}\left(\frac{54321}{73673}\right) &= \left(\frac{73673}{54321}\right) = \left(\frac{19352}{54321}\right) = \left(\frac{2^3\cdot 2419}{54321}\right) = \left(\frac{2419}{54321}\right)\\ &= \left(\frac{54321}{2419}\right) = \left(\frac{1103}{2419}\right) = -\left(\frac{2419}{1103}\right) = -\left(\frac{213}{1103}\right) = -\left(\frac{1103}{213}\right)\\ &= -\left(\frac{38}{213}\right) = -\left(\frac{2\cdot 19}{213}\right) = \left(\frac{19}{213}\right) = \left(\frac{213}{19}\right) = \left(\frac{4}{19}\right) = 1.\end{aligned}$$

So it is a quadratic residue.

Python

The method followed in the above example is very much like the Euclidean algorithm. Here you start with two numbers with greatest common divisor equal to 1 and in every step factors 2 are taken apart and the sign is adjusted.

arithmetics.py

```
def factors2(a):
    v2 = 0
    while a % 2 == 0:
        a, v2 = a // 2, v2 + 1
    return a, v2

def jacobi(a, b):
    p = 0
    a, b = a % b, b
    if a == 0: return 0
    f2 = factors2(a)
    a, b, p = f2[0], b, (f2[1] * ((b * b - 1) // 8) + p) % 2
    while a > 1:
        a, b, p = b % a, a, ((a - 1) * (b - 1) % 8) // 4 + p % 2
        if a == 0: return 0
        else:
            f2 = factors2(a)
            a, b, p = f2[0], b, (f2[1] * ((b * b - 1) // 8) + p)\
            % 2
    return (-1)**p
```

```
>>> jacobi(3, 19)
-1
>>> jacobi(543456543409090, 234452611773869)
1
```

14.7 Square Roots in $\mathbb{F}_p$

If we know that $\overline{a}$ is a square in the field $\mathbb{F}_p$, for example by computing $\left(\frac{a}{p}\right)$, Then that does not mean that we know a square root of $\overline{a}$. In this section it is shown that extracting square roots in $\mathbb{F}_p$ can be done by exponentiation. As we have seen exponentiation in modular arithmetic can be done fast. We start with a simple case.

14.35 Proposition. *Let p be a prime number with $p \equiv 3 \pmod 4$. Let a (with $p \nmid a$) be a quadratic residue modulo p. Then for $b = a^{\frac{p+1}{4}}$ we have $b^2 \equiv a \pmod p$.*

Proof. This follows from Euler's Criterion:

$$b^2 \equiv a^{\frac{p+1}{2}} \equiv a^{\frac{p-1}{2}} a \equiv \left(\frac{a}{p}\right) a \equiv a \pmod p. \qquad \square$$

A bit more complicated case:

14.36 Proposition. *Let p be a prime number with $p \equiv 5 \pmod 8$. Let a (with $p \nmid a$) be a quadratic residue modulo p. Then for $b = a^{\frac{p+3}{8}}$ and $c = 2^{\frac{p-1}{4}} b$ we have $b^2 \equiv a \pmod p$ or $c^2 \equiv a \pmod p$.*

Proof. Let $k = \mathrm{o}_p(a)$ and let $p - 1 = 4t$. Then t is odd. We have $k \mid \frac{p-1}{2}$, say $k = 2^i k_0$ with $k_0 \mid t$ and $i = 0$ or $i = 1$. Then $\gcd(k, \frac{p-1}{4})(= \gcd(k, t))$ is odd. So from proposition 13.33 follows

$$\mathrm{o}_p(a^{\frac{p-1}{4}}) = \frac{k}{\gcd(k,t)} = \frac{2^i k_0}{k_0} = 2^i.$$

For $i = 0$ we have

$$b^2 = a^{\frac{p+3}{4}} = a^{\frac{p-1}{4}} a \equiv a \pmod p$$

and for $i = 1$

$$c^2 = 2^{\frac{p-1}{2}} a^{\frac{p-1}{4}} a \equiv (-1)(-1)a \pmod p. \qquad \square$$

The case $p \equiv 1 \pmod 8$ remains.

14.37 Proposition. *Let p be a prime number with $p \equiv 1 \pmod 8$, say $p - 1 = 2^s t$ with $s \geq 3$ and t odd. Let a (with $p \nmid a$) be a quadratic residue modulo p. Choose an integer d such that $\left(\frac{d}{p}\right) = -1$. Then there is a j with $0 \leq 2j \leq 2^s$ such that $b = a^{\frac{t+1}{2}} d^{t(2^{s-1}-j)}$ satisfies $b^2 \equiv a \pmod p$.*

Proof. From $\mathrm{o}_p(d) \nmid \frac{p-1}{2}(= 2^{s-1}t)$ follows $\mathrm{o}_p(d^t) \nmid 2^{s-1}$. So $\mathrm{o}_p(d^t) = 2^s$, because $\mathrm{o}_p(d^t) \mid 2^s$. We also have $\mathrm{o}_p(a^t) \mid 2^s$. So $\overline{a}^t$ is an even power of $\overline{d}^t$, say $\overline{a}^t = \overline{d}^{t \cdot 2j}$ with j such that $0 \leq 2j \leq 2^s$, that is $0 \leq j \leq 2^{s-1}$. Then

$$a^t \cdot d^{t(2^s - 2j)} \equiv 1 \pmod p,$$

and so

$$a^{t+1} \cdot d^{t(2^s-2j)} \equiv a \pmod{p},$$

that is

$$\left(a^{\frac{t+1}{2}} \cdot d^{t(2^{s-1}-j)}\right)^2 \equiv a \pmod{p}. \qquad \square$$

This proposition is a special case of a result published in 1891 by the Italian mathematician **Alberto Tonelli** (1850-1920). It inspired the American mathematician **Daniel Shanks** (1917-1996) what is now called the Tonelli-Shanks algorithm for finding square roots.

14.38 Example. From $\left(\frac{5}{89}\right) = \left(\frac{89}{5}\right) = \left(\frac{4}{5}\right) = 1$ it follows that 5 is a quadratic residue modulo 89. Now 89 is a small number and the square root of $\overline{5}$ is easily found by trying. However, we apply proposition 14.37. We have $89 - 1 = 88 = 2^3 \cdot 11$. First we compute $\overline{5}^{11}$:

$$\begin{aligned} 5^2 &= 25 \\ 5^4 &= 625 \equiv 2 \pmod{89} \\ 5^5 &\equiv 10 \pmod{89} \\ 5^{10} &\equiv 100 \equiv 11 \pmod{89} \\ 5^{11} &\equiv 55 \pmod{89}. \end{aligned}$$

We look for a d with $\left(\frac{d}{89}\right) = -1$. From $\left(\frac{3}{89}\right) = \left(\frac{89}{3}\right) = \left(\frac{2}{3}\right) = -1$ follows that we can choose $d = 3$. Next we compute $\overline{3}^{11}$:

$$\begin{aligned} 3^2 &= 9 \\ 3^4 &= 81 \\ 3^5 &= 243 \equiv 65 \pmod{89} \\ 3^{10} &\equiv 4225 \equiv 42 \pmod{89} \\ 3^{11} &\equiv 126 \equiv 37 \pmod{89}. \end{aligned}$$

We have $\mathrm{o}_{89}(37) = 8$ and also $\mathrm{o}_{89}(55) \mid 8$. So $\overline{55}$ is a power of $\overline{37}^2 (= \overline{34})$:

$$\begin{aligned} 34^2 &= 1156 \equiv 88 \equiv -1 \pmod{89} \\ 34^3 &\equiv -34 \equiv 55 \pmod{89} \end{aligned}$$

So $37^6 \equiv 55 \pmod{89}$. Now we have $5^{11}37^2 \equiv 1 \pmod{89}$. So

$$5 \equiv 5^{12} \cdot 37^2 \equiv (5^6 \cdot 37)^2 \equiv (50 \cdot 37)^2 \equiv (1850)^2 \equiv 70^2 \pmod{89}.$$

Python

According to the above propositions extraction of square roots is a combination of exponentiation and finding a nonsquare residue (in case of the last proposition). Exponentiation for modular arithmetic we already have. Finding a nonsquare residue is a matter of trying; half of the residue classes modulo p is a nonsquare. We add a function `sqrt(a,p)` which returns to a given a with $\left(\frac{a}{p}\right) = 1$ a b with $b^2 \equiv a \pmod{p}$.

arithmetics.py

```
import random

def sqrt(a, p):
    a = a % p
    p8 = p % 8
    if p8 == 3 or p8 == 7:
       return pow(a, (p + 1) // 4, p)
    if p8 == 5:
        x = pow(a, (p + 3) // 8, p)
        c = pow(x, 2, p)
        if c != a: x = modprod(x, pow(2, (p - 1) // 4, p), p)
        return x
    if p8 == 1:
        d = 3
        while jacobi(d, p)==1:
            d = random.randint(2, p - 1)
        t, s = factors2(p - 1)
        A = pow(a, t, p)
        D = pow(d, t, p)
        m = 0
        for i in range(s):
            if pow(modprod(A, pow(D, m, p), p), 2**(s - 1 - i),
            p) + 1 == p:
                m = m + 2**i
        return modprod(pow(a, (t + 1) // 2, p), pow(D, m // 2, p),
        p)
```

The prime number 345676543456009933 below is found using methods of the next chapter.

```
>>> jacobi(345668887987, 345676543456009933)
1
>>> sqrt(345668887987, 345676543456009933)
144398962591515745
>>> pow(144398962591515745, 2, 345676543456009933)
345668887987
```

14.8 Representation by Quadratic Forms (2)

In section 14.1 we have seen which natural numbers are representable by the form $x^2 + y^2$. We will use the quadratic reciprocity law for the representability by some more quadratic forms. First a useful proposition.

14.39 Proposition. *Let a, m and n be integers with*

a) *a is not a square,*
b) *m or $-m$ is a prime number,*
c) *mn is representable by the form $x^2 - ay^2$,*
d) *m is representable by the form $x^2 - ay^2$.*

Then also n is representable by the form $x^2 - ay^2$.

Proof. There are integers s, t, u, v with $mn = s^2 - at^2$ and $m = u^2 - av^2$. Now to determine $x, y \in \mathbb{Z}$ such that $n = x^2 - ay^2$. If numbers x and y satisfy this identity, then by proposition 14.3

$$mn = (ux + avy)^2 - a(uy + vx)^2.$$

Does the system

$$\begin{aligned} ux + avy &= s \\ vx + uy &= t \end{aligned}$$

of equations have a solution with $x, y \in \mathbb{Z}$? We multiply the first equation by v and the second by u. Subtraction yields $(u^2 - av^2)y = ut - vs$, that is $my = ut - vs$. Next we find $mx = us - avt$. From $mn = s^2 - at^2$ and $m = u^2 - av^2$ follows $s^2 \equiv at^2 \pmod{|m|}$ and $u^2 \equiv av^2 \pmod{|m|}$. Hence

$$u^2t^2 \equiv v^2s^2 \pmod{|m|} \quad \text{and} \quad u^2s^2 \equiv a^2v^2t^2 \pmod{|m|},$$

that is

$$m \mid (ut - vs)(ut + vs) \quad \text{and} \quad m \mid (us - avt)(us + avt).$$

Since $|m|$ is a prime number we have

$$m \mid ut - vs \quad \text{or} \quad m \mid ut + vs$$

and

$$m \mid us - avt \quad \text{or} \quad m \mid us + avt.$$

Replacing u by $-u$ if necessary, makes that we can assume $m \mid ut - vs$.

> Suppose $m \nmid us - avt$. Then $m \mid us + avt$. From $us - avt = us + avt - 2avt$ follows $m \neq \pm 2$. Then also $m \nmid us$ and $m \nmid avt$. So in particular $m \nmid s, v$. Furthermore,
>
> $$(us - avt)(ut + vs) = (u^2 - av^2)ts + (s^2 - at^2)uv = m(ts + nuv),$$

which implies $m \mid ut + vs$ and so $m \mid 2vs$. Since $m \neq \pm 2$ we have $m \mid vs$. Contradiction.

Hence $m \mid us - avt$. So there are $x, y \in \mathbb{Z}$ with $n = x^2 - ay^2$, namely $x = \frac{us-avt}{m}$ and $y = \frac{ut-vs}{m}$. □

14.40 Representation by $x^2 + 2y^2$. The essential prime divisors of the form $x^2 + 2y^2$ are the odd prime numbers p with $\left(\frac{-2}{p}\right) = 1$.

$$\left(\frac{-2}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{2}{p}\right) = (-1)^{\frac{p-1}{2}}(-1)^{\frac{p^2-1}{8}} = (-1)^{\frac{(p-1)(p+5)}{8}}.$$

So the essential prime divisors are the prime numbers p with $p \equiv 1, 3 \pmod 8$.

Let p be a prime number with $p \equiv 1 \pmod 8$ or $p \equiv 3 \pmod 8$. Then there is an $s \in \mathbb{Z}$ with $s^2 \equiv -2 \pmod 8$. From lemma 14.7 follows that there are $x, y \in \mathbb{Z}$ with $(x, y) \neq (0, 0)$, $x^2, y^2 < p$ and $p \mid x^2 + 2y^2$. Then $0 < x^2 + 2y^2 < 3p$ and so $x^2 + 2y^2 = p$ or $x^2 + 2y^2 = 2p$. Since 2 is representable by the form $x^2 + 2y^2$, proposition 14.39 implies that also in the second case p is representable by the form $x^2 + 2y^2$. Analogous to the case $a = -1$ (theorem 14.9) we now have:

Let $n \in \mathbb{N}^+$. Then n is representable by the form $x^2 + 2y^2$ if and only if $v_p(n)$ is even for all prime numbers p with $p \equiv 5, 7 \pmod 8$.

14.41 Representation by $x^2 + 3y^2$. The essential prime divisors of the form $x^2 + 3y^2$ are the prime numbers p with $p \neq 2, 3$ and $\left(\frac{-3}{p}\right) = 1$. From $\left(\frac{-3}{p}\right) = \left(\frac{p}{3}\right)$ these are the prime numbers p with $p \equiv 1 \pmod 3$.

Let p be a prime number with $p \equiv 1 \pmod 3$. Then there is an $s \in \mathbb{Z}$ with $s^2 \equiv -3 \pmod p$. From lemma 14.7 follows that there are $x, y \in \mathbb{Z}$ with $(x, y) \neq (0, 0)$, $x^2, y^2 < p$ and $p \mid x^2 + 3y^2$. Then $0 < x^2 + 3y^2 < 4p$ and so $x^2 + 3y^2 = p$, $x^2 + 3y^2 = 2p$ or $x^2 + 3y^2 = 3p$. Since 3 is representable, in the third case it follows that p is representable. The second case does not occur, for otherwise 2 would be a quadratic residue modulo 3. We have:

Let $n \in \mathbb{N}^+$. Then n is representable by the form $x^2 + 3y^2$ if and only if $v_p(n)$ is even for all prime numbers p with $p \equiv 2 \pmod 3$.

14.42 Representation by $x^2 - 2y^2$. The essential prime divisors of the form $x^2 - 2y^2$ are the odd prime numbers p with $\left(\frac{2}{p}\right) = 1$. These are the prime numbers p with $p \equiv 1, 7 \pmod 8$.

Let p be a prime number with $p \equiv 1, 7 \pmod 8$. Then there is an $s \in \mathbb{Z}$ with $s^2 \equiv 2 \pmod p$. From lemma 14.7 it follows that there are $x, y \in \mathbb{Z}$ with $(x, y) \neq (0, 0)$, $x^2, y^2 < p$ and $p \mid x^2 - 2y^2$. Then $-2p < x^2 - 2y^2 < p$ and so $x^2 - 2y^2 = -p$. Since -1 is representable, p is representable if $-p$ is. So here we have:

Let $n \in \mathbb{Z}$ with $n \neq 0$. Then n is representable by the form $x^2 - 2y^2$ if and only if $v_p(n)$ is even for all prime numbers p with $p \equiv 3, 5 \pmod 8$.

14.43 Representation by $x^2 - 3y^2$. The essential prime divisors of the form $x^2 - 3y^2$ are the prime numbers p with $p \neq 2, 3$ and $\left(\frac{3}{p}\right) = 1$. By example 14.27 these are the prime numbers p with $p \equiv 1, 11 \pmod{12}$.

Let p be a prime number with $p \equiv 1, 11 \pmod{12}$. Then there is an $s \in \mathbb{Z}$ with $s^2 \equiv 3 \pmod{p}$. From lemma 14.7 follows that there are $x, y \in \mathbb{Z}$ with $(x, y) \neq (0, 0)$, $x^2, y^2 < p$ and $p \mid x^2 - 3y^2$. Then $-3p < x^2 - 3y^2 < p$ and so $x^2 - 3y^2 = -p$ or $x^2 - 3y^2 = -2p$. Since -2 is representable, in the second case it follows that p is representable. In the first case we have $p \equiv 2 \pmod{3}$ and so $p \equiv 11 \pmod{12}$. Similarly we get in the first case $p \equiv 1 \pmod{12}$. Here we have:

> *Let $n \in \mathbb{Z}$. Then n is representable by the form $x^2 - 3y^2$ if and only if $v_p(n)$ is even for all prime numbers p with $p \equiv 5, 7 \pmod{12}$ and the sign of n is equal to $(-1)^N$ with*
>
> $$N = v_2(n) + v_3(n) + \sum_{p \equiv 11 \pmod{12}} v_p(n),$$
>
> *that is n is negative if the number of prime factors 2 plus the number of prime factors 3 plus the number of prime factors $\equiv 11 \pmod{12}$ is odd and otherwise n is positive.*

In Chapter 21 we study the Diophantine equations $x^2 - ay^2 = 1$ and $x^2 - ay^2 = -1$ where $a > 0$ and not a square. See also the exercises 14 and 16.

Exercises

1. Let P be a finite nonempty set of prime numbers and let N be the product of these prime numbers: $N = \prod_{p \in P} p$.
   (i) Show that $4N - 1$ has a prime divisor p with $p \equiv 3 \pmod{4}$.
   (ii) Prove that there are infinitely many prime numbers p with $p \equiv 3 \pmod{4}$.
   (iii) Show that $N^2 + 1$ has a prime divisor p with $p \equiv 1 \pmod{4}$.
   (iv) Prove that there are infinitely many prime numbers p with $p \equiv 1 \pmod{4}$.
2. Write $2^3 \cdot 3^4 \cdot 5^5 \cdot 7^6$ as a sum of two squares.
3. Determine all squares in $\mathbb{F}_{17}$, all fourth powers and also all eighth powers. Show that the transformation $f \colon \mathbb{F}_{17} \to \mathbb{F}_{17}$, $x \mapsto x^3$ is bijective. The inverse of f is of type $x \mapsto x^m$ for some $m \in \mathbb{N}$. Which m?
4. Prove that $\overline{\frac{p+1}{2}}$ is a square in $\mathbb{F}_p$ if and only if $\overline{2}$ is so.
5. For which prime numbers p is 5 a square modulo p? And 7?
6. Which prime numbers are essential prime divisors of the form $x^2 + 7y^2$?

7. 123457 is a prime number. Verify that 76775 is a quadratic residue modulo 123457. Do this using Jacobi symbols. Try to do it using Legendre symbols only.

8. Let p be an odd prime number. Let a, b and c be integers with $p \nmid a$. Show that the number of solutions in $\mathbb{F}_p$ of the quadratic equation $\overline{a}x^2 + \overline{b}x + \overline{c} = \overline{0}$ is equal to $\left(\frac{b^2-4ac}{p}\right) + 1$.

9. Let p be a prime number with $p \equiv 3 \pmod 4$ and let $2p+1$ be a prime number also. Verify that $\left(\frac{2}{2p+1}\right) = 1$. Derive from this that $2p+1 \mid 2^p - 1$. (So $2^p - 1$ is not a prime number if $p \equiv 3 \pmod 4$ and $p \geq 7$.)

10. Let p be an odd number. Prove that

$$\left(\frac{2}{p}\right) = \left(\frac{8-p}{p}\right) = \left(\frac{p}{p-8}\right) = \left(\frac{2}{p-8}\right)$$

and derive from this the formula for $\left(\frac{2}{p}\right)$ anew.

11. Let p be an odd prime number. Let g be a primitive root modulo p.
    (i) Prove (without using Legendre symbols) that for $k \in \mathbb{N}$ we have: g^k is a square modulo p if and only if k is even.
    (ii) Prove Euler's Criterion using part (i).

12. Show that 7 is a square modulo the prime numbers 139, 197 and 113. Determine a square root of 7 modulo each of these prime numbers.

13. Show that $2^3 \cdot 5^6 \cdot 17^5$ is representable by the form $x^2 + 2y^2$ and also by the form $x^2 - 2y^2$. Find such representations.

14. (i) Prove that the Diophantine equation $x^2 - 2y^2 = 1$ has infinitely many solutions. (Hint: use the proof of lemma 14.3.)
    (ii) Prove that also the Diophantine equation $x^2 - 2y^2 = -1$ has infinitely many solutions.
    (iii) Let $n \in \mathbb{Z}$ with $n \neq 0$ be representable by the form $x^2 - 2y^2$. Prove that n is representable by this form in infinitely many ways.

15. (i) Show that $5^4 \cdot 19 \cdot 31$ is representable by the form $x^2 + 3y^2$. Find a representation. Are there others (with other $|x|$ and $|y|$)?
    (ii) Which of the numbers $2^3 \cdot 5^4 \cdot 23$ and $-2^3 \cdot 5^4 \cdot 23$ is representable by the form $x^2 - 3y^2$? Find a representation.

16. Prove that the Diophantine equation $x^2 - 3y^2 = 1$ has infinitely many solutions. And the Diophantine equation $x^2 - 3y^2 = -1$?

17. (i) For which prime numbers p is there an $x \in \mathbb{Z}$ such that $p \mid x^2 + x - 2$?
    (ii) For which prime numbers p is there an $x \in \mathbb{Z}$ such that $p \mid x^2 + x + 2$?

18. Let p be an odd prime number of type $x^2 + 2y^2$ with $x, y \in \mathbb{N}$.
    (i) Are there then also $u, v \in \mathbb{N}$ with $u^2 + 2v^2 = 2p$?

(ii) Show that $\left(\frac{-2}{p}\right) = 1$.

(iii) For which odd prime numbers q is $q - 2$ a quadratic residue modulo q?

19. (i) For which prime numbers p is there an $x \in \mathbb{N}$ such that $p \mid x^2 + 7$?

(ii) For which prime numbers p is there an $x \in \mathbb{N}$ such that $p \mid x^2 - 7$?

15 Prime Tests and Factorization

According to the Fundamental Theorem of Arithmetic positive integers have a unique factorization into primes. Such a factorization contains a lot of information. To be able to use a factorization, one needs to find it. For large numbers that is difficult and on this difficulty cryptographic applications are based, see section 15.5. There do exist fast algorithms for detecting whether a number is composite. Three of such tests will be treated in section 15.2. The outcome of these tests can also be that a number is almost certainly a prime number. In section 15.3 we describe a test which does provide a proof that a given number is prime. Since the 70's of the last century better and better algorithms have been found for finding factors. Because explicit factorization is far more difficult than proving that a number is composite, it is clear that there are composite numbers with unknown prime factors. An example of this is $2^{2^{20}}+1$, the 20-th Fermat number. First we describe some simple techniques which are known already for many centuries.

15.1 Basic Techniques

Numbers are usually given in their decimal representation. The last digit shows whether it is divisible by 2 or 5. For divisibility by 3 one can look at the sum of the digits and for divisibility by 11 at their alternate sum, see exercise 7 of chapter 13. That are useful rules for divisibility by very small primes. In general finding factors can be difficult.

We will describe two methods which are very basic. The first is a systematic search for divisors starting at 2. In the second a list of primes is made long enough to contain a prime divisor if the number is composite.

15.1.1 Searching divisors

The most primitive form of looking for divisors of a number n is to divide repeatedly by numbers less n until a remainder 0 occurs. It suffices to try only numbers d with $d^2 \leq n$: to a divisor d with $d^2 > n$ corresponds the divisor $\frac{n}{d} < d$. If no divisor is found, then n is a prime number. After trying 2 and 3 only numbers d with $\gcd(d,6) = 1$ need to be tried. For such d one has $d \equiv 1 \pmod 6$ or $d \equiv 5 \pmod 6$. Thus the sequence to try is $5, 7, 11, 13, 17, 19, 23, 25, \ldots$. This

sequence is constructed by adding alternately 2 and 4. There remain unnecessary tries (division by 25 being the first).

Factorization of numbers is practically impossible when they have only very large prime factors. In modern cryptography the fact we are unable to do such factorizations is actually used. If p and q are prime numbers in the order of magnitude of 10^{100}, then pq is a number in the order of magnitude of 10^{200}. The least proper divisor is p (if $p < q$). To find it by trying in a systematic way around 10^{100} divisions have to be made. Even with 1 billion tries a second this will take more than 10^{83} years. This number of years is in the order of magnitude of the number of elementary particles in the universe.

Python

The function `trial_factors(n, N)` returns the partial prime factorization of `n` obtained by trying only numbers less than N as divisors. If the remaining factor is less than N^2, then the full prime factorization is found. The function `trial_factorization(n)` returns in principle the prime factorization: it is `trial_factors(n,N)` with N large enough.

arithmetics.py

```
def trial_factors(n, N):
    factors = [n]
    while factors[-1] % 2 == 0:
        factors[-1:] = [2, factors[-1] // 2]
    d = 3
    while d**2 <= factors[-1] and d < N:
        while factors[-1] % d == 0:
            factors[-1:] = [d, factors[-1] // d]
        d = d + 2
    if factors[-1] == 1: return factors[:-1]
    return factors

def trial_factorization(n):
    return trial_factors(n, n)
```

```
>>> trial_factorization(364578654877)
[73, 131, 503, 75793]
```

15.1.2 Eratosthenes's sieve

To produce a list of all prime numbers less than a given N proceed as follows: start with the list $2, 3, \ldots, N-1$. Delete all multiples of 2, except 2 itself, $4, 6, \ldots$, go to the next remaining number p, which must be prime since otherwise it would have been deleted, delete all multiples $2p, 3p, 4p, \ldots$, etc. As soon as the next remaining

Eratosthenes (Cyrene, now Libia 276 BC – Alexandria 194 BC)

Most of Eratosthenes' work has been lost. However it has been described by others. Apart from prime numbers he also studied geometry, in particular straight edge and compass construction problems. He is also known because of his estimation of the Earth's diameter.

number p satisfies $p^2 \geq N$ the process can stop. What remains is a list of prime numbers less than N. This procedure is known as *Eratosthenes's sieve*.

Python

Start with a list of 1's of length N. We will produce a list having a 1 on the p-th place for prime numbers p and 0 elsewhere. First we put a 0 on the 0-th and 1st place. Next we apply Eratosthenes's procedure with the variation that we replace a 1 by a 0 instead of deleting the 1. That is what the function `eratosthenes(N)` does. Eratosthenes's sieve is fast, only addition is used, not multiplication. Obviously, for large N much memory is needed. The function `primes(N)` converts the result of `eratosthenes(N)` to a list with the prime numbers less than N.

arithmetics.py

```
def eratosthenes(N):
    lst = N * [1]
    lst[:2] = [0, 0]
    i, j = 2, 4
    while i**2 < N:
        if lst[i] == 1:
            j = i + i
            while j < N:
                lst[j] = 0
                j = j + i
        i = i + 1
    return lst

def primes(N):
    lst = eratosthenes(N)
    return list(filter(lambda i: lst[i] == 1, range(N)))
```

```
>>> eratosthenes(100)
[0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0,
 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0,
 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0,
 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,
 0, 0, 0, 0, 0, 1, 0, 0]
>>> primes(100)
[2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 6
1, 67, 71, 73, 79, 83, 89, 97]
```

By a minor variation a list is obtained having the least prime divisors instead of zeros. This is realized by the function `factor_table(N)`. Such a list can be used to factorize numbers less than N, as is done by the function `table_factorization(n)`.

arithmetics.py

```
def factor_table(N):
    lst = N * [1]
    i, j = 2, 4
    while i**2 < N:
        if lst[i] == 1:
            j = i + i
            while j < N:
                if lst[j] == 1: lst[j]=i
                j = j + i
        i = i + 1
    return lst

tbl=[]

def table_factorization(n):
    if len(tbl) <= n: tbl[:] = factor_table(n + 1)
    factors = []
    while tbl[n] != 1:
        factors.append(tbl[n])
        n = n // tbl[n]
    factors.append(n)
    return factors
```

```
>>> factor_table(100)
[1, 1, 1, 1, 2, 1, 2, 1, 2, 3, 2, 1, 2, 1, 2, 3, 2, 1, 2, 1, 2, 3, 2,
 1, 2, 5, 2, 3, 2, 1, 2, 1, 2, 3, 2, 5, 2, 1, 2, 3, 2, 1, 2, 1, 2, 3,
 2, 1, 2, 7, 2, 3, 2, 1, 2, 5, 2, 3, 2, 1, 2, 1, 2, 3, 2, 5, 2, 1, 2,
 3, 2, 1, 2, 1, 2, 3, 2, 7, 2, 1, 2, 3, 2, 1, 2, 5, 2, 3, 2, 1, 2, 7,
 2, 3, 2, 5, 2, 1, 2, 3]
>>> table_factorization(1874521)
[11, 19, 8969]
```

From tables it is clear that the 'density' of primes is less for larger numbers. The following function by definition gives for every natural number n the number of primes $\leq n$.

15.1 Definition. We define the function $\pi\colon \mathbb{N}^+ \to \mathbb{N}$ by

$$\pi(n) = \#\{\, p \mid p \text{ is a prime number and } p \leq n \,\}.$$

A table of $\pi(n)$ for some n together with $\frac{n}{\pi(n)}$ exact up to 0,1:

n	$\pi(n)$	$\frac{n}{\pi(n)}$
10	4	2,5
10^2	25	4,0
10^3	168	6,0
10^4	1229	8,1
10^5	9592	10,4
10^6	78498	12,7
10^7	664579	15,0
10^8	5761455	17,4
10^9	50847534	19,7
10^{10}	455052511	22,0

This table suggests an estimate for the fraction $\frac{n}{\pi(n)}$. We return to this briefly in chapter 17 on page 374.

15.2 Pseudoprimes

There are theorems about prime numbers, which for composite numbers do not hold in general. A not so interesting example is:

$$p \text{ is a prime number } \implies p = 2 \text{ or } p \text{ is odd},$$

Such a theorem can be turned into a test: verify whether a given integer $n > 1$ satisfies $n = 2$ or n is odd. If n does not pass it is composite. That is for sure. If n does pass the test, then n could be a prime number, but not necessarily so. In this example there are many composite numbers which pass the test, for example all products of two odd prime numbers. In this section we we will use three more interesting theorems.

Robert Daniel Carmichael (Goodwater (Alabama) 1879 – Merriam (Kansas) 1967)

The Carmichael numbers are named after the American mathematician Robert Carmichael. A still unsolved problem is also named after him: *Carmichael's totient function conjecture*. It states that, for every $n \in \mathbb{N}^+$ there is at least one other $m \in \mathbb{N}^+$ such that $\varphi(m) = \varphi(n)$. In 1907 stated as a theorem, but, after he realized that the proof was false, he stated it in 1922 as an open problem.

15.2.1 Fermat pseudoprimes

If p is a prime number, then by Fermat's Little Theorem, see proposition 13.23, for all $a \in \mathbb{Z}$:

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

So we have for all $a, p \in \mathbb{N}$:

$$p \text{ is a prime number } \implies a^p \equiv a \pmod{p}.$$

For each $a \in \mathbb{N}$ we thus have a prime test, that is if p does not pass the test, then p is not prime. The number a is called the *base* of the prime test.

15.2 Definition. A composite number n is called a *Fermat pseudoprime* or also a *pseudoprime* for the base a if $a^n \equiv a \pmod{n}$. Fermat pseudoprimes for the base 2 are called *pseudoprime* for short.

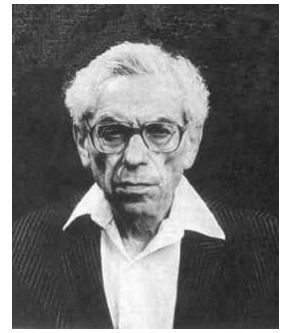
Below 10^{10} there are 455052512 prime numbers and 14884 pseudoprimes. The least pseudoprime is 341 ($= 11 \cdot 31$). This number does not pass the test with base 3. The test can be done for other bases than 2. Annoying is that there are composite numbers that pass the test for all bases: the Carmichael numbers. The smaller the probability a composite number passes the test, the better the test.

15.3 Definition. A natural number which is a Fermat pseudoprime for every base is called a *Carmichael number*.

Carmichael numbers do exist. The least is 561 ($= 3 \cdot 11 \cdot 17$). In 1956 Paul Erdős made it plausible that there are infinitely many Carmichael numbers. A proof was given by Alford, Granville and **Pomerance** in 1994.

Paul Erdős (Budapest 1913 – Warsaw 1996)

The Hungarian mathematician Paul Erdős (Hungarian: Erdős Pál) is by many seen as the founding father of *discrete mathematics*. Solving problems with other mathematicians was his daily activity. He published around 1500 papers, mainly together with other mathematicians, in total more than 500 of them. This resulted in the creation of the *Erdős number*, the length of the shortest path between a mathematician and Erdős in the graph of all mathematicians with co-authorships as the edges.



William Alford (1937-2003) was an American lawyer and a mathematician working in topology and number theory. **Andrew Granville** (1962) is a British number theorist.

15.4 Example. We show that 561 is a Carmichael number. By the Chinese remainder theorem it suffices to show that for all $a \in \mathbb{N}$ we have: $a^{561} \equiv a \pmod 3$, $a^{561} \equiv a \pmod{11}$ and $a^{561} \equiv a \pmod{17}$. This follows directly from Fermat's Little Theorem for the prime numbers 3, 11 and 17.

The following proposition gives necessary and sufficient conditions for a number to be a Carmichael number.

15.5 Proposition. *Let $n \in \mathbb{N}^+$ be composite. Then the following are equivalent:*

(i) *n is a Carmichael number.*
(ii) *$a^{n-1} \equiv 1 \pmod n$ for all $a \in \mathbb{Z}$ with $\gcd(a, n) = 1$.*
(iii) *n is squarefree and $p - 1 \mid n - 1$ for all prime divisors of n.*

Proof.

(i)⇒(ii): Let n be a Carmichael number and let $a \in \mathbb{Z}$ with $\gcd(a, n) = 1$. Then $\overline{a} \in (\mathbb{Z}/n)^*$, so from $\overline{a}^n = \overline{a}$ (in $(\mathbb{Z}/n)^*$) it follows that $\overline{a}^{n-1} = \overline{1}$.

(ii)⇒(iii): Let g be a primitive root modulo a prime divisor p of n (see theorem 13.39). By the Chinese Remainder Theorem we can take g such that $g \equiv 1 \pmod q$ for all prime divisors q of n with $q \neq p$. Then $\gcd(g, n) = 1$ and so $g^{n-1} \equiv 1 \pmod p$. Hence $p - 1 = \mathrm{o}_p(g) \mid n - 1$. By exercise 19 of chapter 13 the integer g can be assumed to be a primitive root modulo p^2. If $p^2 \mid n$, then it would follow that $p(p-1) = \mathrm{o}_{p^2}(g) \mid n - 1$ and so $p \mid n - 1$, contradictory to $p \mid n$. It follows that n is squarefree.

(iii)⇒(i): For each prime divisor p of n we have $a^p \equiv a \pmod p$ for all $a \in \mathbb{Z}$. From this follows that $a^n = a^{n-p}a^p \equiv a^{n-p}a = a^{n-(p-1)} \equiv a^{n-2(p-1)} \equiv$

$\cdots \equiv a \pmod{p}$. By the Chinese Remainder Theorem it follows that $a^n \equiv a \pmod{n}$ for all $a \in \mathbb{Z}$. □

The equivalence (i) $\iff$ (iii) is known as *Korselt's criterium* after the German mathematician **Alwin Korselt** (1864 – 1947). It was a first step in his attempt to show that numbers now known as Carmichael numbers did not exist.

Python

Using the code of the previous chapter the prime test based on Fermat's Little Theorem is easily described.

arithmetics.py

```
def fermat(p, a):
    return pow(a, p, p) == a
```

```
>>> fermat(7463, 2)
False
>>> fermat(341, 2)
True
>>> fermat(341, 3)
False
>>> fermat(561, 2)
True
>>> fermat(561, 17)
True
```

15.2.2 Euler pseudoprimes

Carmichael numbers pass prime tests based on Fermat's Little Theorem. Now we consider prime tests based on Euler's Criterion: for $a \in \mathbb{Z}$, $p \in \mathbb{N}^+$ odd and $\gcd(a, p) = 1$ we have

$$p \text{ is a prime number } \implies a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p}.$$

Here $\left(\frac{a}{p}\right)$ is the Jacobi symbol. We will see that there are no composite numbers which pass the test for all bases a. So in this case there is no analogue of the Carmichael numbers.

15.6 Definition. An odd composite number n is called a *Euler pseudoprime* for the base a if $\gcd(a, n) = 1$ and $a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n}$.

The notion of Euler pseudoprime is a refinement of the notion of Fermat pseudoprime:

15.7 Lemma. *Let n be an odd composite number and $a \in \mathbb{Z}$ with $\gcd(a,n) = 1$. If n is an Euler pseudoprime for the base a, then n is a Fermat pseudoprime for the base a.*

Proof. Since $\gcd(a,n) = 1$, and so $\left(\frac{a}{n}\right) = \pm 1$, the lemma follows from:

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n} \implies a^{n-1} \equiv 1 \pmod{n}. \qquad \square$$

15.8 Example. We will show that the Carmichael number 561 is not an Euler pseudoprime for the base 5. We have $5^{280} \equiv 5^8 \equiv \left(\frac{5}{17}\right) \equiv -1 \pmod{17}$ and therefore not $5^{280} \equiv 1 \pmod{561}$, while $\left(\frac{5}{561}\right) = \left(\frac{5}{3}\right)\left(\frac{5}{11}\right)\left(\frac{5}{17}\right) = (-1)\cdot 1 \cdot (-1) = 1$. In fact $5^{280} \equiv 67 \pmod{561}$.

15.9 Proposition. *Let n be an odd composite number. Then there exists an a with $\gcd(a,n) = 1$ such that n is not an Euler pseudoprime for the base a.*

Proof.

Suppose n is an Euler pseudoprime for all bases a with $\gcd(a,n) = 1$. Then n is a Fermat pseudoprime for all these bases (lemma 15.7). By proposition 15.5 the number n is a Carmichael number. In particular n is squarefree and odd (see exercise 4). Put $n = pm$ with p a prime number. Then $\gcd(p,m) = 1$. Take a $b \in \mathbb{Z}$ with $\left(\frac{b}{p}\right) = -1$. By the Chinese Remainder Theorem there is a $c \in \mathbb{Z}$ such that $c \equiv b \pmod{p}$ and $c \equiv 1 \pmod{m}$. Then $\left(\frac{c}{n}\right) = \left(\frac{c}{p}\right)\left(\frac{c}{m}\right) = \left(\frac{b}{p}\right)\left(\frac{1}{m}\right) = -1 \cdot 1 = -1$. From $c^{\frac{n-1}{2}} \equiv \left(\frac{c}{n}\right) \equiv -1 \pmod{n}$ follows that $c^{\frac{n-1}{2}} \equiv -1 \pmod{m}$. However, $c^{\frac{n-1}{2}} \equiv 1 \pmod{m}$, because $c \equiv 1 \pmod{m}$. Contradiction.

So there is a base for which n is not an Euler pseudoprime. $\square$

If there exists a base for which n is not an Euler pseudoprime, then there are more:

15.10 Proposition. *Let n be an odd composite number. Then there are at least $\frac{\varphi(n)}{2}$ residue classes $\overline{a} \in (\mathbb{Z}/n)^*$ such that n is not an Euler pseudoprime for the base a.*

Proof. Let H be the set of all residue classes $\overline{b} \in (\mathbb{Z}/n)^*$ such that n is an Euler pseudoprime for the base b. Let n be not an Euler pseudoprime for a base a. By proposition 15.9 such an a exists. Multiply the elements of H by $\overline{a}$. Since multiplication by $\overline{a}$ is a permutation of $(\mathbb{Z}/m)^*$, we obtain this way $\#(H)$ residue classes $\overline{ab}$ having the property that n is not an Euler pseudoprime for the base ab. So there are at most $\frac{\varphi(n)}{2}$ residue classes $\overline{a} \in (\mathbb{Z}/n)^*$ such that n is an Euler pseudoprime for the base a. $\square$

15.11 The Solovay-Strassen test. For odd n and a with $\gcd(a,n) = 1$ the remainder of $a^{\frac{n-1}{2}}$ after division by n is easily computed: it is exponentiation in modular arithmetic. Also for the Jacobi symbol $\left(\frac{a}{n}\right)$ the algorithm is fast. It is easily checked whether n is an Euler pseudoprime for the base a. If n passes this test for many randomly chosen bases a, then it is almost sure that the number n is prime. The probability that a composite n passes this test for 100 randomly chosen bases is less than $\frac{1}{2^{100}}$.

The Solovay–Strassen primality test was developed in 1977 by the American mathematician **Robert M. Solovay** and the German mathematician **Volker Strassen**. Solovay (1938) is working in set theory, Strassen (1936) in the analysis of algorithms.

Python

In the test `euler(p, a)` the result of `pow(a, (p - 1) / 2, p)` is compared to that of `jacobi(a, p)`. By `solovay(p, N)` the test `euler(p, a)` is done for N randomly chosen a with $1 < a < p - 2$.

arithmetics.py

```
def euler(p, a):
    j = (jacobi(a, p) % p)
    return  j != 0  and  pow(a, (p - 1) // 2, p) == j

def solovay(p, N):
    for i in range(N):
        a = random.randint(2, p - 1)
        if not euler(p, a): return False
    return True
```

```
>>> euler(561, 2)
True
>>> euler(561, 3)
False
>>> euler(8719309, 2)
False
>>> euler(8719309, 3)
True
>>> len(list(filter(lambda x: euler(561, x), range(1, 561))))
80
>>> len(list(filter(lambda x: euler(8719309, x), range(1, 8719309))))
985608
```

The Carmichael number 561 is for 80 of the $\varphi(561) = 320$ bases an Euler pseudoprime. The Carmichael number 8719309 is for 985608 of the $\varphi(8710309) = 7884864$ bases an Euler pseudoprime; that is a ratio of 1 to 8 bases.

```
>>> solovay(2347, 20)
True
>>> trial_factorization(2347)
[2347]
>>> trial_factorization(239)
[239]
>>> 2347 * 239
560933
>>> len(list(filter(lambda x: euler(560933, x), range(1, 560933))))
578
>>> 2346 * 238
558348
>>> 558348 // 578
966
```

The composite number 560933 is not a Carmichael number; for 1 in 966 bases it is an Euler pseudoprime.

15.2.3 Strong pseudoprimes

We give another refinement of the test based on Fermat's Little Theorem. The basic idea for this test is: if for $\overline{a} \in \mathbb{F}_p$ it holds that $\overline{a}^2 = \overline{1}$, then for $\overline{a}$ there are only two possibilities: $\overline{a} = \overline{1}$ and $\overline{a} = -\overline{1}$. This is so because $\mathbb{F}_p$ is a field.

15.12 Theorem. *Let p be an odd prime number. Write $p - 1 = 2^s t$ with $s \in \mathbb{N}^+$ and t odd. Then for all $a \in \mathbb{Z}$ with $p \nmid a$:*

$$\begin{aligned} a^t &\equiv 1 \pmod{p} \\ \textit{or}\quad a^t &\equiv -1 \pmod{p} \\ \textit{or}\quad a^{2t} &\equiv -1 \pmod{p} \\ \textit{or}\quad a^{2^2 t} &\equiv -1 \pmod{p} \\ &\vdots \\ \textit{or}\quad a^{2^{s-1} t} &\equiv -1 \pmod{p}. \end{aligned}$$

Proof. Consider the sequence

$$\overline{a}^t, \overline{a}^{2t}, \overline{a}^{2^2 t}, \ldots, \overline{a}^{2^{s-1} t}, \overline{a}^{2^s t}$$

of elements of $\mathbb{F}_p^*$. By Fermat's Little Theorem the last element equals $\overline{1}$. So there is a first element in this sequence which is equal to $\overline{1}$. If that is not the first one, then it is preceded by an element $\overline{b}$ with $\overline{b}^2 = \overline{1}$ and $\overline{b} \neq \overline{1}$. Since $\mathbb{F}_p$ is a field, we have $\overline{b} = -\overline{1}$. □

15.13 Definition. Let n be an odd composite number and let $n = 2^s t + 1$ with $s, t \in \mathbb{N}^+$ and t odd. Then n is called a *strong pseudoprime* for the base a if $a^t \equiv 1 \pmod{n}$ or $a^{2^j t} \equiv -1 \pmod{n}$ for some j with $0 \leq j < s$. A strong pseudoprime for the base 2 is also called a strong pseudoprime for short.

For an odd $n \in \mathbb{N}$ we write $n - 1 = 2^s t$ with $s \in \mathbb{N}^+$ and t odd. For $a \in \mathbb{N}$ with $1 < a < n-1$ we can compute $\overline{a}^t$ and if necessary, if not $\overline{a}^t \equiv \pm 1$, square repeatedly. If $-\overline{1}$ occurs, then n passes the test.

Python

We add code for the test based on theorem 15.12.

arithmetics.py

```
def strong(p, a):
    fact = factors2(p - 1)
    b = pow(a, fact[0], p)
    if b == 1 or b == p - 1: return True
    j = 1
    while j < fact[1]:
        b = pow(b, 2, p)
        if b == p - 1: return True
        j = j + 1
    return False
```

```
>>> strong(561, 2)
False
>>> strong(8719309, 2)
False
>>> strong(8719309, 3)
True
```

15.14 Proposition. *Let n be a strong pseudoprime for the base a. Then n is a Fermat pseudoprime for the base a.*

Proof. For each of the cases n passes the test we have $a^{n-1} \equiv 1 \pmod{n}$. □

In exercise 10 it is asked to prove that there are infinitely many strong pseudoprimes. The least strong pseudoprime is 2047. The least number which is a strong pseudoprime for both the base 2 and the base 3 is 1373653. The least that moreover passes the test for the base 5 is 25326001. And when also the base 7 is used it is 3215031751 and below $25 \cdot 10^9$ there are no others.

If n is an composite number, then n passes the test for at most a quarter of the bases a with $1 \leq a < n$. A proof is given in [7]. It is elementary but a bit too technical to present it here.

The question is: 'With how many bases to perform the test to be sure that a number is prime?' The American computer scientist **Gary Miller** has found a bound based on a conjecture in number theory, the generalized Riemann hypothesis. However, nowadays there are faster tests which are not based on conjectures. If the test is performed for k bases, then the probability that a composite number passes the test is less than $\frac{1}{4^k}$. For $k = 100$ the probability is less than 10^{60}, and this is extremely small. If the k bases are chosen at random, then the test is known as the *Miller-Rabin probabilistic primality test*, named after Gary Miller and the Israeli mathematician and computer scientist **Michael Rabin** (1931). If an odd number passes this Miller-Rabin test, then that number is unlikely to be composite. Such a number sometimes is called a *commercial* prime, since for applications as in cryptography absolute certainty is not really necessary: if such a prime behaves as expected everybody is satisfied and it is even more likely that the number is prime.

Python

With `rabin(p, N)` the Miller-Rabin test is applied N times to p, every time with a randomly chosen base.

arithmetics.py

```
def rabin(p, N):
    for i in range(N):
        a = random.randint(2, p - 1)
        if not strong(p, a): return False
    return True
```

```
>>> len(list(filter(lambda x: strong(561, x), range(1, 561))))
10
>>> len(list(filter(lambda x: strong(8719309, x), range(1, 8719309)))
)
246402
>>> len(list(filter(lambda x: strong(560933, x), range(1, 560933))))
578
>>> rabin(43567280831, 1)
False
>>> rabin(43567280837, 1)
True
>>> rabin(43567280837, 100)
True
```

The test can be used to determine the next prime number with high probability, and also to distil primes out of a segment of natural numbers. Here this is done by the functions `next_rabin_prime` and `rabin_prob_primes`.

arithmetics.py

```
def next_rabin_prime(a, N):
    a = a + (a % 2 == 0) + 2 * (a % 2 == 1)
    while not rabin(a, N):
        a = a + 2
    return a

def rabin_prob_primes(a, k, N):
    b = a + k
    a = a + (a % 2 == 0)
    primes = []
    while a < b:
        if a % 3 == 0 or a % 5 == 0: a = a + 2
        else:
          if rabin(a, N):
            primes.append(a)
          a = a + 2
    return primes
```

```
>>> next_rabin_prime(10000000000000000000000000000000, 20)
10000000000000000000000000000049
>>> rabin(10000000000000000000000000000049, 1000)
True
>>> rabin_prob_primes(44444444444444444, 150, 20)
[44444444444444469, 44444444444444557, 44444444444444593]
>>> rabin_prob_primes(555555555555555555, 150, 20)
[555555555555555559, 555555555555555683, 555555555555555691]
```

15.3 The $n-1$-Test

For 'small' $n \in \mathbb{N}^+$ we can find a prime factorization in a systematic way by searching prime divisors or using a factor table. For 'large' numbers we can establish that they are composite or, if we can not, they are probably prime. We want more:

a) for large numbers a way to prove that they are prime,
b) of large numbers known to be composite find a proper factorization.

In this section a method of Lucas is treated which can be used in some cases to prove that a number is prime without trying to factorize the number. This method works best if a prime factorization of the number $n-1$ can be found. Even a partial factorization of $n-1$ might be helpful.

If there is an $a \in \mathbb{Z}/n^*$ with $\mathrm{o}_n(a) = n-1$, then $n-1$ elements in $\mathbb{Z}/n$ are invertible and so only $\overline{0}$ is not. That means that n is a prime number. From lemma 13.41 follows:

15.15 Theorem (Lucas). *Let $n \in \mathbb{N}^+$ and $a \in \mathbb{Z}$ with $\gcd(a,n) = 1$. Suppose that $a^{n-1} \equiv 1 \pmod{n}$ and $a^{\frac{n-1}{q}} \not\equiv 1 \pmod{n}$ for all prime divisors q of $n-1$. Then n is a prime number.* □

Python

The function `primitive(a, n, plist)` tests whether the integer a is of order $n-1$ modulo the integer n. It uses `plist`, a list of all prime divisors of $n-1$. If the function `lucas(p, plist, N)` returns `True`, then an integer of order $p-1$ modulo p has been found, which implies that p is a prime number. The function calls `primitive` at most N times.

arithmetics.py

```
def primitive(a, n, plist):
    if pow(a, n - 1, n) != 1: return False
    for q in plist:
        if pow(a, (n - 1) // q, n) == 1: return False
    return True

def lucas(p, plist, N):
    for i in range(N):
        a = random.randint(2, p - 1)
        if primitive(a, p, plist): return True
    return False
```

```
>>> trial_factorization(6256814489)
[6256814489]
>>> trial_factorization(6256814488)
[2, 2, 2, 4783, 163517]
>>> primitive(2, 6256814489, [2, 4783, 163517])
False
>>> primitive(3, 6256814489, [2, 4783, 163517])
True
>>> next_rabin_prime(842638469359595967887765433,20)
842638469359595967887765513
>>> trial_factorization(842638469359595967887765512)
[2, 2, 2, 199, 3593, 807907, 182339026020661]
>>> lucas(842638469359595967887765513, [2, 199, 3593, 807907, 1823390
26020661], 20)
True
```

The number 6256814489 is prime and the factorization of 6256814488 is used to show that 2 is not a primitive root modulo 6256814489, but that 3 is.

The number 842638469359595967887765513 almost certainly is a prime. The function `lucas` proves that it is a prime.

Fermat primes

Extreme examples of numbers n for which the prime factorization of $n-1$ is no problem are the Fermat numbers.

15.16 Definition. The m-th *Fermat number* F_m is the number $2^{2^m}+1$. If F_m is prime, then F_m is called a *Fermat prime*.

The first five Fermat numbers are 3, 5, 17, 257 and 65537. These are primes indeed. Fermat conjectured that all Fermat numbers are prime. About one century later Euler had shown that $F_5 (= 4294967296)$ is composite: it is divisible by 641. He had no problem in finding this divisor 641, since he understood that prime divisors were of type $64k+1$. This follows from:

15.17 Proposition (Euler). *Let $m \in \mathbb{N}$ and let p be a prime divisor of F_m. Then $p \equiv 1 \pmod{2^{m+1}}$.*

Proof. We have: $2^{2^m} \equiv -1 \pmod{p}$. From this it follows that the order of $\overline{2}$ in $(\mathbb{F}_p)^*$ equals 2^{m+1}. So: $2^{m+1} \mid p-1$. □

Lucas sharpened this result somewhat:

15.18 Proposition (Lucas). *Let $m \in \mathbb{N}$ with $m \geq 2$ and let p be a prime divisor of F_m. Then $p \equiv 1 \pmod{2^{m+2}}$.*

Proof. From $m \geq 2$ it follows that $2^3 \mid p-1$, that is $p \equiv 1 \pmod 8$. By theorem 14.24 we have that 2 is a square modulo p, and so $2^{m+1} \mid \frac{p-1}{2}$. □

> The Fermat numbers F_m with $5 \leq m \leq 32$ are composite. For not a single $m \geq 12$ the prime factorization of F_m is known, though for some of these Fermat numbers a proper factorization has been found. It is unknown whether there is yet another Fermat prime. For m equal to 20 and to 24 no prime divisor is known. The largest Fermat number known to be composite is $F_{18233954}$: a prime factor $7 \cdot 2^{18233956}+1$ was found in 2020.

Pépin's test is a prime test for Fermat numbers. The test is named after the French theologist and mathematician **Jean François Theophile Pépin** (1826–1904).

15.19 Theorem (Pépin). *For $m \in \mathbb{N}^+$:*

$$F_m \text{ is prime} \iff 3^{\frac{F_m-1}{2}} \equiv -1 \pmod{F_m}.$$

Proof.

⇒: Let F_m be prime. Then by Euler's Criterion $3^{\frac{F_m-1}{2}} \equiv \left(\frac{3}{F_m}\right) \pmod{F_m}$. We compute $\left(\frac{3}{F_m}\right)$ using quadratic reciprocity:

$$\left(\frac{3}{F_m}\right) = \left(\frac{F_m}{3}\right) = \left(\frac{(-1)^{2^m}+1}{3}\right) = \left(\frac{2}{3}\right) = -1.$$

$\Leftarrow$: If $3^{\frac{F_m-1}{2}} \equiv -1 \pmod{F_m}$, then $3^{F_m-1} \equiv 1 \pmod{F_m}$. Since 2 is the unique prime divisor of $F_m - 1$, F_m is prime by theorem 15.15. □

The greatest Fermat number for which Pépin's test is used successfully is F_{24}.

Python

The code for Pépin's test is simple:

arithmetics.py

```
def pepin(m):
    F = (2**(2**m)) + 1
    return pow(3, (F - 1) // 2, F) == F - 1
```

`pepin(m)` tests the m-th Fermat number for being prime.

```
>>> pepin(3)
True
>>> pepin(4)
True
>>> pepin(5)
False
>>> pepin(6)
False
>>> pepin(14)
False
>>> trial_factorization(2**(2**5) + 1)
[641, 6700417]
>>> trial_factorization(2**(2**6) + 1)
[274177, 67280421310721]
```

Using a partial factorization of $n-1$

Let n be an odd number which is probably prime. A prime factorization of $n-1$ can be used to prove that n is prime. Even a partial factorization of $n-1$ is useful.

15.20 Theorem (Pocklington). *Let n be an odd number and $n-1 = ku$ with $k, u \in \mathbb{N}^+$. Let $a \in \mathbb{Z}$ with $a^{n-1} \equiv 1 \pmod{n}$ and $\gcd(a^{\frac{n-1}{q}} - 1, n) = 1$ for all prime divisors q of k. Then $p \equiv 1 \pmod{k}$ for every prime divisor p of n.*

Proof. Let p be a prime divisor of n. Then $(a^u)^k = a^{n-1} \equiv 1 \pmod{p}$ and so $\mathrm{o}_p(a^u) \mid k$. Moreover, $p \nmid a^{\frac{n-1}{q}} - 1 = (a^u)^{\frac{k}{q}} - 1$ for all prime divisors q of k and so by lemma 13.41 we have $\mathrm{o}_p(a^u) = k$. It follows that $k \mid p-1$, that is $p \equiv 1 \pmod{k}$. □

15.21 Corollary. *If moreover, in the context of theorem 15.20, $k^2 \geq n$, then n is prime.*

Proof. Let p be any prime divisor of n. Then $p \equiv 1 \pmod{k}$ and so $p^2 > k^2 \geq n$. So n is prime. □

An $n-1$-test can be based on Corollary 15.21.

Python

The function `pocklington(n, k, K)` determines for an odd n with a divisor k of $n-1$, and where K is the list of prime divisors of k, whether prime divisors of n are congruent 1 modulo k.

arithmetics.py

```
def pocklington(n, k, plist, N):
    for i in range(N):
        a = random.randint(2, n - 1)
        if primitive(a, n, plist): return True
    return False
```

```
>>> next_rabin_prime(236455865876488352477, 20)
23645586587648835248 3
>>> trial_factorization(23645586587648835248 2)
[2, 101, 2791, 388793, 1078749107]
>>> pocklington(23645586587648835248 3, 23645586587648835248 2, [2, 101
, 2791, 388793, 1078749107])
True
```

The number $n = 23645586587648835248 3$ is probably a prime number. A factorization of $n-1$ was easily found and it was used for showing that indeed n is prime. In other cases more steps might be necessary:

```
>>> next_rabin_prime(2364558658764883525050, 20)
2364558658764883525103
>>> trial_factors(2364558658764883525102, 1000000)
[2, 29, 40768252737325578019]
>>> rabin(40768252737325578019, 20)
True
>>> trial_factors(4076825273732557801 8, 1000000)
[2, 3, 3, 2264902929851421001]
>>> trial_factorization(226490292985142100 0)
[2, 2, 2, 3, 5, 5, 5, 59, 12796061750573]
>>> pocklington(226490292985142100 1, 12796061750573, [12796061750573]
)
True
>>> pocklington(4076825273732557801 9, 226490292985142100 1, [226490292
9851421001])
True
```

```
>>> pocklington(236455865876488352510 3, 40768252737325578019, [407682
52737325578019])
True
```

15.4 Factorization

If a number is known to be composite, then it is very well possible that a factorization is unknown. When a straightforward search for a divisor is not successful another method might be used. One of such methods is known as Pollard-rho, introduced in 1975 by the British mathematician **John Pollard** (1941). Later other methods were designed, but here we confine to Pollard-rho.

15.4.1 The Pollard-rho factorization algorithm

We will use a transformation f of $\mathbb{Z}$ which satisfies the following property:

$$x \equiv y \pmod{m} \implies f(x) \equiv f(y) \pmod{m} \quad \text{for all } m \in \mathbb{N}^+ \text{ and } x, y \in \mathbb{Z}.$$

This means that f induces a transformation $\overline{f}$ of $\mathbb{Z}/m$ for all $m \in \mathbb{N}^+$. For example any map of type $x \mapsto x^2 + c$, where $c \in \mathbb{Z}$, satisfies this condition. Let $a \in \mathbb{Z}$ and consider the course of a under f:

$$a, f(a), f^2(a), f^3(a), \ldots$$

Since $\mathbb{Z}/m$ is finite, for every m the course

$$\overline{a}, \overline{f(a)}, \overline{f^2(a)}, \ldots$$

of $\overline{a} \in \mathbb{Z}/m$ under $\overline{f}$ repeats, say $\overline{f^r(a)} = \overline{f^s(a)}$ for some $r > s$. Then $\overline{f^{k+(r-s)}(a)} = \overline{f^k(a)}$ for all $k \geq s$. Draw a picture of this sequence and of the action of $\overline{f}$ on its terms and it is clear why 'rho' occurs in the algorithm's name.

Let n be a composite number. For the computation of the numbers $\gcd(f^i(a) - f^j(a), n)$ only the remainders of $f^i(a)$ after division by n are needed. As soon as r and s are found with $1 < \gcd(f^r(a) - f^s(a), n) < n$, one has a proper divisor of n, namely the number $\gcd(f^r(a) - f^s(a), n)$.

> Let p be a prime divisor of n. After how many terms of the sequence can one expect that $p \mid f^r(a) - f^s(a)$? How many terms have to be computed to have two terms which are congruent modulo p? This is related to the birthday problem, see also 11.4. The expected number is in the order of magnitude of $\sqrt{p}$.

Any a may be chosen as starting value and also any c (if the transformation is of type $x \mapsto x^2 + c$) may be chosen at random. Since the success of the algorithm depends on these choices, one calls such a method of computing a *Monte-Carlo method*.

Algorithm

There are tricks to fasten the algorithm by reducing the number of times $\gcd(f^i(a) - f^j(a), n)$ have to be computed. There are ways not to keep the numbers $f^i(a)$, thus reducing the required memory. A nice way is the *Floyd cycle finding method*. In this method only the differences $f^{2j}(a) - f^j(a)$ are computed: the difference j of $2j$ and j will be for a certain j a multiple of the length of the period, while at the same time $f^j(a)$ is in the period.

Python

Pollard-rho using the Floyd cycle finding method is in Python easily described. The starting value a and the c in the function $x \mapsto x^2 + c$ are chosen at random, so equal inputs may have different results.

arithmetics.py

```
def pollardrho(n):
    g = n
    while g == n:
        c = random.randint(1, n - 3)
        a = random.randint(0, n - 1)
        u = v = a
        def F(x):
            return (pow(x, 2, n) + c) % n
        g = 1
        while g == 1:
            u, v = F(u), F(F(v))
            g = gcd(u - v, n)
    return g
```

```
>>> p = next_rabin_prime(2637897656751, 20)
>>> p
2637897656761
>>> q = next_rabin_prime(2675634056751, 20)
>>> q
2675634056807
>>> n = p * q
>>> n
7058048808801113661622127
>>> pollardrho(n)
2675634056807
>>> trial_factorization(53751794982079)
[26539, 32467, 62383]
>>> pollardrho(53751794982079)
32467
>>> pollardrho(53751794982079)
26539
```

15.4.2 Prime factorizations

We have:

a) An algorithm for finding prime factors up to a given magnitude.
b) An algorithm for finding a proper divisor of a composite number: Pollard-rho.
c) Algorithms to determine with high probability that an odd number is prime: the Miller-Rabin and Solovay-Strassen prime tests.
d) An algorithm for proving a number to be prime: the $n-1$-test.

In principle the first algorithm suffices for the factorization. The problem is that it might take far too much time to end. For 'large' numbers n we can make use of the other algorithms. For example:

1. Use algorithm 1 for finding the prime factors less than 10^9.
2. If a factor greater than 10^{18} remains, use one of the algorithms 3 to determine whether this factor is composite or probably prime.
3. If the factor is probably prime, then try with algorithm 4 to prove that it actually is prime. For this $n-1$-test (partial) prime factorization of $n-1$ is needed and for that do this factorization process for $n-1$.
4. If the factor is composite, then use algorithm 2 to find a factorization of this factor. For the factors found do this factorization process.

It can happen that a lot of book keeping is needed. That can be automated, but here we will not do so. We do by hand using the computer, interactively so to say.

Python

We take an integer n.

```
>>> n = 216536215346334573546647504549540054776 63
>>> trial_factors(n, 10000000)
[216536215346334573546647504549540054776 63]
```

There are no prime divisors less than 10000000.

```
>>> rabin(n, 20)
False
```

So n is composite. With `pollardrho` we find a divisor.

```
>>> pollardrho(n)
22746427603
>>> a = n // 22746427603
>>> a
95195702430994422534002296602 1
>>> rabin(a, 20)
True
```

A divisor 22746427603 has been found. It is prime, since otherwise there is a prime divisor < 10000000 and such a divisor would already have been found. The other factor, the number a, is probably prime. We will prove it is prime using the $n-1$-test.

```
>>> trial_factors(a - 1, 10000000)
[2, 2, 3, 5, 29, 43, 293, 43424219232412361925277]
>>> b = 4342421923241236192527 7
>>> rabin(b, 20)
True
```

The number b is probably prime and again we use the $n-1$-test.

```
>>> trial_factors(b - 1, 10000000)
[2, 2, 3, 189619, 6939103, 2750208289]
>>> pocklington(b, b - 1, [2, 2, 3, 189619, 6939103, 2750208289], 5)
True
>>> pocklington(a, b, [b], 5)
True
```

The divisors of $b-1$ which have been found are small enough to conclude that they are prime. From the $n-1$-test follows that b is prime. Another application of the $n-1$-test shows that a is prime. So we have the prime factorization of n:

$$n = 22746427603 \cdot 95195702430994422534002296602 1.$$

In this example we were lucky to find a factor with Pollard-rho. This factor is relatively small and that makes it so that the algorithm did not take long to find it. We have already seen that F_7, the seventh Fermat number is composite. Again we find a factor using `pollardrho`, though it takes some time on an ordinary desktop computer.

```
>>> def Fermat(n):
...     return 2**(2**n) + 1
...
>>> Fermat(7)
340282366920938463463374607431768211457
>>> pollardrho(Fermat(7))
59649589127497217
>>> d = Fermat(7) // 59649589127497217
>>> d
5704689200685129054721
>>> rabin(d, 20)
True
>>> trial_factorization(d - 1))
[2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 5, 12497, 733803839347]
>>> lucas(d, [2, 3, 5, 12497, 733803839347], 20)
True
```

The prime factorization of F_7, known since 1970:

$$F_7 = 59649589127497217 \cdot 5704689200685129054721.$$

Next we try F_8.

```
>>> Fermat(8)
115792089237316195423570985008687907853269984665640564039457584007913
129639937
>>> pollardrho(Fermat(8))
1238926361552897
>>> k = Fermat(8) // 1238926361552897
>>> k
93461639715357977769163558199606896584051237541638188580280321
>>> rabin(k, 20)
True
>>> factors = trial_factors(k - 1, 100000000)
>>> factors
[2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 5, 7,13,33432649280047353540365856
1554225677455540427332439719911]
>>> rabin(factors[-1],20)
False
>>> pollardrho(factors[-1])
31618624099079
>>> l = factors[-1] // 31618624099079
>>> l
10573720467811625362740343546868933296253329
>>> rabin(l, 20)
True
>>> lucas(k, [2, 3, 5, 7, 13, 31618624099079, l], 20)
True
```

The prime factorization of F_8, known since 1980:

$$F_8 = 1238926361552897 \cdot 93461639715357977769163558199606896584051237541638188580280321.$$

15.5 RSA Cryptosystems

A simple secret code can be made by taking a permutation of the 26 letters of the alphabet. Such a code is given by a list of pairs, for example

a	b	c	d	e	f	g	h	i	j	k	l	m	⋯
x	c	k	g	h	o	f	u	q	b	y	v	w	⋯

Coding a text is done by applying the permutation to each of the letters in the text. The result can be decoded by applying the inverse permutation. If the permutation is as simple as that, decoding is only a little bit harder than coding the text. For decoding the letters have to be looked for in the bottom row, where they are not in the usual order. If the code is not known, decoding is more difficult, but in this case it is still easy, especially when a coded text of some length is available.

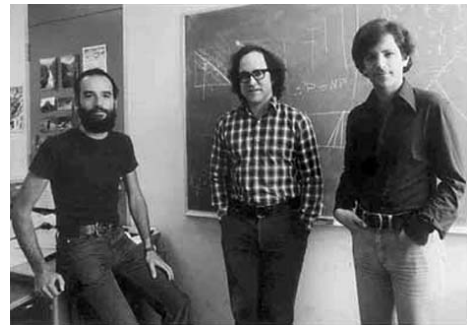
In this section we will describe a code which is practically impossible to break, even if the code itself is public, that is it is not secret. This is not as strange as it may seem. For example if one permutes all words of the English language instead of only permuting the 26 letters of the alphabet. Such a permutation may be given by a dictionary. Then it is hard to find a given decoded word. But nowadays such a job is easily done by a computer. We will describe a permutation of something like 10^{200} objects. Decoding is then practically impossible unless there is enough extra information.

All characters we might use in a text can be replaced by numbers in a standard way. We could use for example ASCII-code. In the binary notation 8 digits per character are used (8 binary digits = 8 bits = 1 byte). Thus a text can be converted in a standard way into a (large) number by concatenation of all bytes. Coding in this way will come down to applying a permutation to the set of all numbers less than a given large number. This number might be in the order of magnitude of 10^{200}. To put it differently, coding will be a permutation of all possible 'texts' of length comparable to the length of a single line in a book.

First we describe the code. It depends on two large prime numbers p and q, both in the order of magnitude of 10^{100}. There are plenty of these primes and they are easily found by computer, more so if you are not after absolute certainty. Rabin's test does this job very well. These prime numbers we keep secret, not their product m. The code is a permutation of $\mathbb{Z}/m$. A system of representatives is $\mathbb{N}_m$, the set of all natural numbers less than m. We choose a number e such that $\gcd(e, \varphi(m)) = 1$. We know that $\varphi(m) = (p-1)(q-1)$, but others do not, since they are not able to factorize m. The code is the following permutation of $\mathbb{Z}/m$:

$$\mu_e \colon \mathbb{Z}/m \to \mathbb{Z}/m, \qquad x \mapsto x^e,$$

Leonard Adleman (San Francisco 1945)



In 1976 Adleman became assistant professor at the MIT (Cambridge, Massachusetts, USA). His colleague **Ronald Rivest** (1947) of Computer Science had the idea to look for a code with a public key. Adleman and Rivest's colleague **Adi Shamir** (1952) became interested. Adleman's role mainly was cracking other's codes. He succeeded in the first 42 cases, but not so in the 43rd. That code is now known as the RSA-code.

Adleman considered his contribution as unimportant and, moreover, less interesting than his other work. So in his opinion there was no need to name him as one of the authors. The others thought differently. He agreed to be the last in the list. That explains the order of the letters RSA. On the photograph from left to right Shamir, Rivest and Adleman. Later, in the 90's, Adleman became well-known for his work on DNA computers.

(raising to the power e). This transformation is a permutation because it has an inverse: since $\gcd(e, \varphi(m)) = 1$, there exist $f, n \in \mathbb{Z}$ such that $ef + \varphi(m)n = 1$ and μ_f (raising to the power f) is the inverse: for $x \in (\mathbb{Z}/m)^*$ we have

$$\mu_f \mu_e(x) = \mu_f(x^e) = (x^e)^f = x^{ef} = x^{ef} x^{\varphi(m)n} = x^{ef+\varphi(m)n} = x,$$

and using the Chinese Remainder Theorem it is not hard to see that even $\mu_f \mu_e(x) = x$ for all $x \in \mathbb{Z}/m$ (and not only in $(\mathbb{Z}/m)^*$). So decoding is raising to the power f. Knowing p and q, implies knowing $\varphi(m)$ and then f is easily found using the extended Euclidean algorithm. If only m and e are known, then coding is possible, but for decoding the only thing to do is to factorize m, and with the present state of knowledge and hardware that will probably take thousands of years. The code described here is the *RSA code*, named after its inventors: Rivest, Shamir and Adleman. The success of the code rests on one hand on our knowledge (the recognition of large primes and the ability to do modular arithmetic fast), and on the other hand our lack of knowledge (we are not able to factorize large numbers).

A simple RSA-crypto system with Python

The ASCII-code of the characters on a keyboard varies from 32 up to 126. Subtracting 32 each of the characters is represented by a two digits. We will use this for the translation of `strings` into `lists` of digits and backwards.

Strings of length l are translated into lists of digits of length $2l$. We determine random primes p and q of length (= length of the decimal representation) l such that their product $m = pq$ is (not much) greater then 10^{2l}. Next we determine a random e satisfying $\gcd(e, (p-1)(q-1)) = 1$ and by the extended Euclidean algorithm the f needed for decoding. We add the function `makersa(l)`.

arithmetics.py

```
def makersa(l):
    p = next_rabin_prime(random.randint(10**(l - 1), 9 * \
    (10**(l - 1))), 20)
    q = next_rabin_prime(random.randint((10**(2 * l)) // p,
    (10**(2 * l)) // p + (10**(l - 1))), 20)
    m = p * q
    e = random.randint(10**(2 * l - 1), 9 * (10**(2 * l - 1)))
    while gcd(e, (p - 1) * (q - 1)) > 1:
        e = e + 1
    f = modinv(e, (p - 1) * (q - 1))
    return (l, m, e), (l, m, f), (p, q)
```

The function `makersa(l)` returns $((l, m, e), (l, m, f), p, q)$. The first component (l, m, e) is a code that can be used for coding strings of length l. The second component (l, m, f) is for decoding. Or the other way round if so inclined. The numbers p and q are returned as well, but are not needed for the crypto system.

Here we work with character strings of arbitrary length without carriage return. They will be chopped into strings of length l. Refinements of the procedure are possible, but now it is only the principle that matters.

Coding of a `string` `s` is as follows:

1. Convert s into a `list` of characters.
2. Convert this list into a list of numbers: take the `ascii`-code minus 32.
3. Concatenate the numbers to obtain one large number.
4. Chop this into numbers of length l (with zeros added to the last number if necessary).
5. Transform each of the numbers using the code `(l,m,e)`.

The result is a list of numbers.

arithmetics.py
```
def transform(nrlist, a, n):
    return [pow(nr, a, n) for nr in nrlist]

def encode_rsa(s, code):
    def number(c): return ord(c) - 32
    def str2(n): return str(n).zfill(2)
    def codenumber(s):
        return int(''.join(map(str2, map(number, list(s)))))
    def codenrs(lst): return [codenumber(s) for s in lst]
    return transform(codenrs([s[code[0] * i:code[0] * (i + 1)]
    .ljust(code[0]) for i in range(len(s) // code[0] + 1)]),
    code[2], code[1])
```

Decoding of the `list` of numbers `nrlst` is done in the opposite order:

1. Transform each of the numbers using the code `(l,m,f)`.
2. Concatenate the numbers to obtain one large number.
3. Chop this number into a list of numbers of 2 digits.
4. Replace each of the numbers in the list by the character having that number plus 32 as `ascii`-code.
5. Convert the list of characters into one `string`.

The result is the original `string s` (with a number of spaces added at the end as a result of chopping into words of equal length).

arithmetics.py
```
def decode_rsa(nrlst, code):
    def char(n):
        if n > 94: return r' '
        else: return chr(n + 32)
    def phrase(codenr, N):
        nrstr = str(codenr).zfill(N)
        return ''.join(map(char, map(int, [nrstr[2 * i:2 * i + 2]
        for i in range(len(nrstr) // 2)])))
    def phrases(nrlst, N):
        return [phrase(nr, N) for nr in nrlst]
    return r''.join(phrases(transform(nrlst, code[2], code[1]),
    2*code[0]))
```

If a code is made using prime numbers of length 15, then it is easily broken, in the next example in just a few minutes with the use of Pollard-rho. Such primes are too small.

```
>>> rsa = makersa(15)
>>> rsa
((15, 10767934257520169348175343940 39, 751537927195162228107676912199
), (15, 10767934257520169348175343940 39, 428170688654659395393699451487
), (847620309727837, 1270372374746147))
>>> pollardrho(10767934257520169348175343940 39)
1270372374746147
```

Let us make a code using prime numbers of length 100.

```
>>> rsa = makersa(100)
>>> rsa
((100, 10106065487380965708910022486588869680541130468954619460542501
63408682298333301941737509379420890936182532636423395493192199985928835
46545071897429085300077755142548753937878801165628600817767665185717,
 81419591363844748860494932152623479050294745287675211646294327820447
07235887792205862541722900587099182220065326105872502783532042854998111
96371852491126983759023982363369721954550831323036497394545 63), (10
85935492217793758679315729746283251613097616063523436082796883646454723
31660973490201315789952093 9, 93063221156364109663570393565845763569
04102606847728730481416164584643207594271149336311375564781110 3))
```

It works:

```
>>> encmesg = encode_rsa(r"Since we are not able to factorize numbers
 of 200 digits, the RSA-code made using primes of 100 digits is a saf
 e public code", rsa[0])
>>> encmesg
[13553281190459677236091651158840559043483367252510877369153790518119
17901236293985243523419077400744441020335426144278331703379738193337 6
81125531797215938845445431854819369565875847747652955393023006 5, 9542
03191879789958974379282401631980580022054668477952144365901714741601 1
48628192831740306415210420932404898305102289334564565016921042249355 1
94441676007924232676307983926995039881981929756873952453 76]
>>> decode_rsa(encmesg, rsa[1])
'Since we are not able to factorize numbers of 200 digits, the RSA-co
de made using primes of 100 digits is a safe public code
                                                                    '
```

Exercises

1. Factorize 11601 by looking for divisors.
2. Make a table of the least prime divisors of the composite numbers < 170. Make this table in the same way as Eratosthenes' sieve is made. How can the prime factorizations of the numbers 91, 97 and 117 be determined using this table?
3. Verify that 561, 6601 and 8719309 ($= 19 \cdot 37 \cdot 79 \cdot 157$) are Carmichael numbers.
4. Can a Carmichael number be even?
5. (i) Let d and n be natural numbers with $d \mid n$. Prove that $2^d - 1 \mid 2^n - 1$.
   (ii) Let n be an odd pseudoprime. Prove that $2^n - 1$ is such as well.
   (iii) Show that there are infinitely many pseudoprimes.
6. Show that 561 is an Euler pseudoprime for the base 2 and that 121 an Euler pseudoprime for the base 3.
7. Let n be an odd pseudoprime.
   (i) Prove that $2^n - 1$ is an Euler pseudoprime for the base 2.
   (ii) Show that there are infinitely many Euler pseudoprimes for the base 2.
8. Let n be odd and $a \in \mathbb{Z}$ with $\gcd(a, n) = 1$. Prove that the following are equivalent:

   n is a Fermat pseudoprime for the base a.

   n is a Euler pseudoprime for the base a^2.
9. Show that all odd composite numbers are Euler pseudoprimes for the base -1.
10. (i) Let n be an odd pseudoprime. Prove that $2^n - 1$ is a strong pseudoprime.
    (ii) Show that there are infinitely many strong pseudoprimes.
11. (i) Prove that for all $m \in \mathbb{N}^+$ we have $F_m = F_0 \cdots F_{m-1} + 2$.
    (ii) Prove that $\gcd(F_i, F_j) = 1$ if $i \neq j$.
    (iii) Show that from (ii) it follows that there are infinitely many prime numbers.
12. Prove using the $n - 1$-test that 139309 is a prime number.
13. (i) We have $2^{1150} \equiv 1 \pmod{1151}$ and $2^{230} \equiv 1060 \pmod{1151}$. Does it follow with Pocklington's method that 1151 is a prime number?
    (ii) The number 1169 is small and has a small prime divisor. If we apply Pollard-rho to 1169 using the transformation $x \mapsto x^2 + 1$ and start value 1, a proper divisor is readily found. Which one? How fast?
    (iii) Is 34 a square modulo 1151 ?
    (iv) Is 34 a square modulo 1169 ?
14. We consider the natural numbers $H_m = \frac{3^{2^m}+1}{2}$, where $m \in \mathbb{N}^+$.

(i) Show that H_m is odd for all $m \in \mathbb{N}^+$.

(ii) Let $m \in \mathbb{N}^+$. Show that $\overline{3} \in \mathbb{Z}/{H_m}^*$ and determine the order of 3 modulo H_m.

(iii) Let $m \in \mathbb{N}^+$ and let p be a prime divisor of H_m. Prove that $p \equiv 1 \pmod{2^{m+1}}$.

(iv) Factorize H_3 using part (iii).

15. Prove the Theorem of Proth:
Let $n = 2^s t + 1$ with $s, t \in \mathbb{N}^+$ and $2^s > t$. If there is an $a \in \mathbb{Z}$ such that $a^{\frac{n-1}{2}} \equiv -1 \pmod{n}$, then n is a prime number.

16. Let $n = q^s t + 1$ with q a prime number, $s, t \in \mathbb{N}^+$ and $q^s > t$. Prove: if there is an $a \in \mathbb{Z}$ satisfying $a^{n-1} \equiv 1 \pmod{n}$ and $a^{\frac{n-1}{q}} \not\equiv 1 \pmod{n}$, then n is a prime number.

17. Factorize 8633 using Pollard-rho.

18. Let $n = pq$ with p and q different prime numbers. Prove that $a^{\varphi(n)+1} \equiv a \pmod{n}$ for all $a \in \mathbb{Z}$.

19. Show that each element of $\mathbb{Z}/341$ is a 7th power.

20. 641 is a prime number. Show that the map

$$\mathbb{F}_{641} \to \mathbb{F}_{641}, \quad x \mapsto x^{427}$$

is the inverse of the map

$$\mathbb{F}_{641} \to \mathbb{F}_{641}, \quad x \mapsto x^3.$$

21. Let n a be Carmichael number and let $n = pm$ with p a prime number and $m \in \mathbb{N}^+$. The numbers p and n satisfy $p \equiv n \equiv 5 \pmod{8}$. Let $a \in \mathbb{N}^+$ with $a \equiv 2 \pmod{p}$ and $a \equiv 1 \pmod{m}$.

(i) Compute $\left(\frac{a}{n}\right)$.

(ii) Prove that n is not an Euler pseudoprime for the base a.

(iii) Show that $n - 1 = c(p - 1)$ for an odd $c \in \mathbb{N}^+$.

(iv) Prove that $a^{\frac{n-1}{2}} \not\equiv \pm 1 \pmod{n}$.

(v) Prove that n is not a strong pseudo-prime for for the base a. (Example: $n = 8719309$ and $p = 37$.)

Part IV

Completions

In chapter 9 the field $\mathbb{Q}$ of rational numbers has been constructed. That was a temporary endpoint in the construction of the number system. There are two reasons to want more. The first is an algebraic one: not all polynomial equations have solutions. Fields can be extended in such a way that polynomials do have zeros. In chapter 20 we will do this in simple cases: we adjoin square roots. This is also done in chapter 19 in a special, but important, case: the field $\mathbb{C}$ of complex numbers is constructed by adjoining the square root of -1 is to the field $\mathbb{R}$ of real numbers, the field we will construct in chapter 17. The second reason is an analytic one: there exist sequences of numbers which seem to tend to a number, but the number is so far not available. An example of a number we would like to have is π, the ratio of the circumference of a circle to its diameter. Here it is not clear whether this ratio is rational, it is even not made precise what actually is meant by this ratio. Chapter 16 is about limits, what it means that a sequence of numbers approaches a number, but also about sequences that should approach a number though in $\mathbb{Q}$ it does not exist.
For defining limits a notion of distance between numbers is needed. For that purpose absolute values are introduced. An absolute value tells what the distance to 0 is. On $\mathbb{Q}$ we have the ordinary absolute value. In chapter 17 the field $\mathbb{R}$ of real numbers is constructed, numbers which can be approximated by rational numbers using this notion of distance.
On $\mathbb{Q}$ there are more absolute values. We will have a look at these as well. They result in another notion of distance and so in another notion of limit, also limits outside the field $\mathbb{Q}$. In fact for each prime p we have an absolute value on $\mathbb{Q}$. In chapter 18 we will for each prime p extend $\mathbb{Q}$ to the field $\mathbb{Q}_p$ of p-adic numbers, which also consists of limits of sequences in $\mathbb{Q}$. Chapter 20 contains an example of the use of p-adic numbers in number theory.

For the construction of the field $\mathbb{C}$ only the field $\mathbb{R}$ is needed, so for this construction one can skip chapter 18.

16 Limits

In this chapter two types of absolute values are defined: the ordinary absolute value in section 16.1 and the p-adic absolute value (defined in section 16.6), depending on a given prime p. The first one everybody is familiar with, but for an understanding of the second it usually takes some time. Up to the last section only the ordinary absolute value is considered. In the sections 16.2 and 16.3 limits of sequences or rational numbers are considered. In section 16.4 the g-adic notation for rational numbers is studied (for $g = 10$ it is the familiar decimal notation). Most interesting are the Cauchy sequences of rational numbers (section 16.5). We use them in the next chapter for the construction of the field of real numbers. In the last section it is done all over for the p-adic absolute value.

> *In this chapter all numbers are rational numbers. So far these are all the numbers we have!*

16.1 The Ordinary Absolute Value on $\mathbb{Q}$

In subsection 7.4.6 the absolute value of integers was introduced. It can easily be extended to the rational numbers.

16.1 Definition. Let $r \in \mathbb{Q}$. The *absolute value* $|r|$ of r is defined by

$$|r| = \begin{cases} r & \text{if } r \geq 0, \\ -r & \text{if } r \leq 0. \end{cases}$$

Thus we have a map $\mathbb{Q} \to \mathbb{Q}^{\geq 0}, r \mapsto |r|$. This map is called the *(ordinary) absolute value on* $\mathbb{Q}$.

> It is called ordinary because it is the absolute value which is used in most of mathematics. For the other absolute values see the last section.

The properties of the absolute values of integers proven in proposition 7.34 hold for the absolute values of rational numbers as well:

16.2 Proposition. *The absolute value* $\mathbb{Q} \to \mathbb{Q}^{\geq 0}$, $r \mapsto |r|$ *has the following properties:*

(i) $|r| = 0 \iff r = 0$ *(for all* $a \in \mathbb{Q}$*),*
(ii) $|rs| = |r| \cdot |s|$ *(for all* $r, s \in \mathbb{Q}$*),*
(iii) $|r + s| \leq |r| + |s|$ *(for all* $r, s \in \mathbb{Q}$*).*

Proof. See the proof of proposition 7.34. □

The absolute value of a number can be seen as the distance of that number to number 0. The distance between two numbers should be preserved under addition of any number to these numbers, or as one might say, the distance is invariant under translation.

16.3 Definition. Let r and s be rational numbers. The *distance* $d(r, s)$ of r to s is the absolute value of the difference of r and s:

$$d(r, s) = |r - s|.$$

16.4 Proposition. *The distance* $d\colon \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}^{\geq 0}$ *of rational numbers is a* metric *on* $\mathbb{Q}$*, that is for all* r*,* s *and* t *in* $\mathbb{Q}$*:*

(i) $d(r, s) = 0 \iff r = s$,
(ii) $d(r, s) = d(s, r)$,
(iii) $d(r, t) \leq d(r, s) + d(s, t)$ (the triangle inequality).

Proof. The triangle inequality follows from Proposition 16.2(iii):
$|(r - s) - (s - t)| \leq |r - s| + |s - t|$. □

Approximations with decimal fractions

We are used to expressions like 24.8045. These refer to special rational numbers: $24.8045 = \frac{248045}{10000}$. Often such a number is intended to be an approximation for the exact number:

> r equals 24.8045 up to 4 decimals

usually means $24.80445 \leq r < 24.80455$, that is $\frac{2480445}{100000} \leq r < \frac{2480455}{100000}$. Then the distance of r to $\frac{248045}{10000}$ is less than $\frac{1}{20000}$. Another way of rounding off is made by simply deleting all further decimals: then the distance of r to $\frac{248045}{10000}$ is less than $\frac{1}{10000}$.

Let r be a rational number and $n \in \mathbb{N}$. Then $0 \leq 10^n r - \lfloor 10^n r \rfloor < 1$ and so

$$0 \leq r - \frac{\lfloor 10^n r \rfloor}{10^n} < \frac{1}{10^n}.$$

16.5 Definition. Rational numbers of type $\dfrac{a}{10^n}$ with $a \in \mathbb{Z}$ and $n \in \mathbb{N}$ are called *decimal fractions*.

Let r be a rational number. By taking n large enough there is a decimal fraction on a distance of r less than any given positive number ε: there is an $N \in \mathbb{N}$ with $10^N > \frac{1}{\varepsilon}$ and for every $n \geq N$ we have

$$0 \leq r - \frac{\lfloor 10^n r \rfloor}{10^n} < \frac{1}{10^n} \leq \frac{1}{10^N} < \varepsilon.$$

No matter how small ε, there is a decimal fraction on a distance less than ε.

16.2 Null Sequences

A sequence $a_0, a_1, a_2, \ldots$ of numbers we will often denote by (a_n). The parentheses are an indication that we are dealing with a sequence. For the numbering the index n is used. Usually we start with index 0, sometimes with 1 and it could also be another integer. Since we use sequences for approximating numbers, we are not really interested in the first terms of the sequence. If we nevertheless want to indicate what the first index is, then we might use a notation like $(a_n)_{n\geq 1}$.

16.6 Definition. A sequence (a_n) of rational numbers is called a *null sequence* if for every $\varepsilon > 0$ an $N \in \mathbb{N}$ exists such that

$$|a_n| < \varepsilon \quad \text{for all } n \geq N.$$

The definition of null sequence expresses what it means that the terms of a sequence approach 0, or in the terminology of the next section: 0 is the limit of the sequence. For centuries mathematicians used sequences for approaching a number while exact definitions were still lacking. In the nineteenth century Cauchy made all this much more precise and Weierstraß gave definitions as we use them nowadays.

The definition requires the existence of an N for every $\varepsilon > 0$. Note that if an N satisfies the requirement for a given ε, every natural number greater than N satisfies this requirement as well.

The numbers ε in the definition are rational: at this moment all our numbers are rational. Later we consider more generally sequences of real numbers and then we also admit real numbers ε, though, as we will see, it will not make any difference for the notion of null sequence.

16.7 Example. The sequence (a_n) with $a_n = \frac{1}{n}$ is a null sequence. We give a detailed proof. Let ε be any positive number. Take $N = \lfloor \frac{1}{\varepsilon} \rfloor + 1$. Then $N > \frac{1}{\varepsilon}$ and so for all $n \geq N$:

$$|a_n| = |\tfrac{1}{n}| = \tfrac{1}{n} \leq \tfrac{1}{N} < \varepsilon.$$

So for every $\varepsilon > 0$ there is an N with the property that $|a_n| < \varepsilon$ for all $n \geq N$.

Augustin-Louis Cauchy (Paris 1789 – Sceaux 1857)

The French mathematician Cauchy made a start with the exact formulation of the notion of limit. He worked among other subjects on complex functions and determinants. His 789 articles are brought together in the 27 volumes of his collected works. He was a conservative Roman Catholic who was involved in many disputes and conflicts.

In this example the number N comes out of the blue. Such an N is often found by reasoning backwards: for which n do we have $\frac{1}{n} < \varepsilon$, that is $n > \frac{1}{\varepsilon}$? This clearly holds for all natural numbers greater than the floor of $\frac{1}{\varepsilon}$.

The definition of null sequence is quite subtle. The difficulty lies primarily in the alternation 'all, there is an, all' in the definition. This alternation is not unusual for notions where some kind of approaching is involved. Some consider it instructive to formulate this as a game. Every sequence (a_n) determines a (very short) game for two players. These players make the following moves:

1. Player 1 gives an $\varepsilon > 0$.
2. Player 2 gives an $N \in \mathbb{N}$.
3. Player 1 gives an $n \geq N$.

If $|a_n| < \varepsilon$, then player 2 wins. If $|a_n| \geq \varepsilon$, then player 1 wins. If there exists a winning strategy for player 2, then (a_n) is a null sequence. If there exists a winning strategy for player 1, then (a_n) is not a null sequence .

In the example there is a winning strategy for player 2: give an N with $N > \lfloor \frac{1}{\varepsilon} \rfloor + 1$.

Another formulation of the definition of null sequence is as follows:

Definition. A sequence is a *null sequence* if for every $\varepsilon > 0$ the ε-neighborhood of 0 contains almost all terms of the sequence.

Thus 'there exists an $N \in \mathbb{N}$' and 'all $n \geq N$' are hidden in 'almost all' and 'ε-neighborhood'. By *almost all* we understand: all but a finite number. The *ε-neighborhood* of 0 consists of all numbers having absolute value less than ε.

16.8 Definition. Let (a_n) be a sequence and $(i(n))$ a sequence in $\mathbb{N}$ with $i(0) < i(1) < i(2) < \cdots$. Then the sequence $(a_{i(n)})$ is called a *subsequence* of the sequence (a_n).

In particular $(i(n))$ is a subsequence of (n). From $i(0) < i(1) < i(2) < \cdots$ follows easily (by mathematical induction) that $i(n) \geq n$ for all $n \in \mathbb{N}$.

Karl Theodor Wilhelm Weierstraß (Ostenfelde 1815 – Berlin 1897)

The notion of limit as we use it today comes from the German mathematician Karl Weierstraß. He made important contributions to mathematical analysis, especially to the theory of complex functions. Health problems caused him to lecture sitting down, while a student wrote for him on the blackboard.

16.9 Lemma. *Let the sequence (a_n) of rational numbers be a null sequence. Then every subsequence of (a_n) is a null sequence as well.*

Proof. Let $(a_{i(n)})$ be a subsequence and let $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ such that $|a_n| < \varepsilon$ for all $n \geq N$. For $n \geq N$ we then have $i(n) \geq n \geq N$ and so $|a_{i(n)}| < \varepsilon$. □

16.10 Example. The sequence (a_n) with $a_n = \frac{1}{10^n}$ is a subsequence of the sequence $(\frac{1}{n})$ and so is a null sequence as well. So we have

$$|a_n| = |\tfrac{1}{10^n}| \leq \tfrac{1}{n} < \varepsilon$$

for any $\varepsilon > 0$ if $n \geq \lfloor \frac{1}{\varepsilon} \rfloor + 1$. Here too $N = \lfloor \frac{1}{\varepsilon} \rfloor + 1$ satisfies. Clearly, N could have been chosen much smaller, but here our only concern is the existence of such an N.

16.11 Lemma. *Let (a_n) be a sequence of rational numbers. Then*

$$(a_n) \text{ is a null sequence} \iff (|a_n|) \text{ is a null sequence .}$$

Proof. This is an immediate consequence of the definition: whether (a_n) is a null sequence, only depends on the absolute values $|a_n|$. □

16.12 Example. The sequence (a_n) with $a_n = \frac{(-1)^n}{n}$ is a null sequence. The terms of this sequence are alternately less and greater than 0.

16.13 Lemma. *Let (a_n) and (b_n) be sequences of rational numbers. Suppose (b_n) is a null sequence and for all n we have $|a_n| \leq b_n$. Then (a_n) is a null sequence as well.*

> Apparently the sequence (b_n) has only nonnegative terms. The condition in this lemma is also phrased as: the sequence (a_n) is *bounded above* by the null sequence (b_n).

Proof. Let $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ such that $b_n = |b_n| < \varepsilon$ for all $n \geq N$. For these n we also have $|a_n| < \varepsilon$. □

The sequence $(\frac{1}{10^n})$ being a null sequence, is a special case of (a^n) being a null sequence for any a such that $|a| < 1$. We will prove this using the Bernoulli inequality:

16.14 Proposition (Bernoulli inequality). *Let x be a rational number with $x \geq -1$. Then for all $n \in \mathbb{N}$:*

$$(1+x)^n \geq 1 + nx.$$

Proof. We use mathematical induction. For $n = 0$ it is clear. If it holds for some $n \in \mathbb{N}$, then also for its successor:

$$\begin{aligned}(1+x)^{n+1} &= (1+x)^n(1+x)\\ &\geq (1+nx)(1+x) && \text{(because } 1+x \geq 0)\\ &= 1 + (n+1)x + nx^2 \geq 1 + (n+1)x\end{aligned}$$

□

16.15 Proposition. *Let $a \in \mathbb{Q}$ with $|a| < 1$. Then the sequence (a^n) is a null sequence.*

Proof. We assume that $a \neq 0$. Then $\frac{1}{|a|} > 1$ and so $\frac{1}{|a|} - 1 > 0$. By the Bernoulli inequality

$$\frac{1}{|a|^n} = \Big(1 + \frac{1}{|a|} - 1\Big)^n \geq 1 + n\Big(\frac{1}{|a|} - 1\Big) > n\Big(\frac{1}{|a|} - 1\Big),$$

that is

$$|a^n| < \frac{|a|}{1-|a|} \cdot \frac{1}{n}.$$

Since $(\frac{|a|}{1-|a|} \cdot \frac{1}{n})$ is a null sequence, also (a^n) is a null sequence (lemma 16.13). □

The sum of two null sequences is a null sequence:

16.16 Proposition. *Let (a_n) and (b_n) be null sequences. Then also $(a_n + b_n)$ is a null sequence.*

Proof. Let $\varepsilon > 0$. There is an $M \in \mathbb{N}$ such that $|a_n| < \frac{\varepsilon}{2}$ for all $n \geq M$ and there is an $N \in \mathbb{N}$ such that $|b_n| < \frac{\varepsilon}{2}$ for all $n \geq N$. For $n \geq \max(M, N)$ we then have

$$|a_n + b_n| \leq |a_n| + |b_n| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon.$$

□

Also the product of two null sequences is a null sequence, but not necessarily both sequences are null sequences, we can do with less, see proposition 16.19.

16.17 Definition. A number sequence (a_n) is called *bounded* if there exists a number C such that $|a_n| \leq C$ for all $n \in \mathbb{N}$.

Often we will conclude that a sequence is bounded if there exists a C with $|a_n| \leq C$ for all n from a certain N onwards: take the constant $\max(|a_0|, \ldots |a_{N-1}|, C)$.

16.18 Lemma. *Null sequences are bounded.*

Proof. Let $(a_n)_{n\geq 0}$ be a null sequence. Take in the definition of null sequence $\varepsilon = 1$: there is an $N \in \mathbb{N}$ with $|a_n| < 1$ for all $n \geq N$. So $|a_n| \leq \max(1, |a_0|, \ldots, |a_{N-1}|)$ for all $n \in \mathbb{N}$. □

16.19 Proposition. *Let the sequence (a_n) be bounded and the sequence (b_n) a null sequence. Then also $(a_n b_n)$ is a null sequence.*

Proof. There is a C such that $|a_n| \leq C$ for all $n \in \mathbb{N}$. Then $|a_n b_n| \leq C \cdot |b_n|$. By the lemmas 16.11 and 16.13 we have that $(a_n b_n)$ is a null sequence. □

16.3 Convergent Sequences

In this section all sequences are sequences of rational numbers.

16.20 Definition. Let (a_n) be a sequence of rational numbers. We say that (a_n) *converges* to a rational number a if the sequence $(a_n - a)$ is a null sequence. The number a is called the *limit* of the sequence (a_n). Notation: $\lim_n a_n = a$. If a sequence converges to a number, we say that the sequence *converges* or that it is *convergent*. We also express this by saying that $\lim_n a_n$ exists.

The sequence $(a_n - a)$ being a null sequence means that for each $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $|a_n - a| < \varepsilon$ for all $n \geq N$, in other words for each $\varepsilon > 0$ almost all terms lie in the ε-neighborhood of a. That neighborhood consists of all numbers at a distance from a less than ε.

A sequence can converge to only one number. That is why we can speak of *the* limit of a convergent sequence: if both $(a_n - a)$ and $(a_n - b)$ are null sequences, then is so is their difference, being the constant sequence $(b - a)$, that is $b = a$.

16.21 Proposition. *Let (a_n) be a sequence converging to a and (b_n) a sequence converging to b. Then the sequence $(a_n + b_n)$ converges to $a + b$.*

Proof. The sequence $(a_n + b_n - a - b)$ is a null sequence, since it is the sum of the null sequences $(a_n - a)$ and $(b_n - b)$, see proposition 16.16. □

We also formulate this as the rule

$$\lim_n (a_n + b_n) = \lim_n a_n + \lim_n b_n.$$

In this rule the limit of the sum of two sequences is expressed in terms of the limits of the sequences *if* these limits do exist. In formulas like this one it is usually understood that the limits on the right hand side exist.

16.22 Proposition. *Let (a_n) be a convergent sequence. Then the difference sequence $(a_{n+1} - a_n)$ is a null sequence.*

Proof. The sequence (a_{n+1}) is a subsequence of (a_n). Both have the same limit. From proposition 16.21 it follows that the difference sequence converges to 0. □

Later we will see examples of sequences having a null sequence as difference sequence and which nevertheless do not converge.

16.23 Lemma. *Convergent sequences are bounded.*

Proof. If (a_n) converges to a, then $(a_n - a)$ is a null sequence and is therefore bounded by lemma 16.18: there is a C such that $|a_n - a| \leq C$ for all n. Then $|a_n| \leq |a| + |a_n - a| \leq |a| + C$ for all n and so (a_n) is bounded as well. □

16.24 Proposition. *Let (a_n) be a sequence converging to a and (b_n) a sequence converging to b. Then the sequence $(a_n b_n)$ converges to ab.*

Proof. We prove that the sequence $a_n b_n - ab$ is a null sequence:

$$a_n b_n - ab = a_n b_n - a_n b + a_n b - ab = a_n(b_n - b) + (a_n - a)b.$$

The sequences $(b_n - b)$ and $(a_n - a)$ are null sequences. The sequence (a_n) is bounded (lemma 16.23) and so is the constant sequence (b). The proposition now follows from the propositions 16.19 and 16.21. □

So the rule is

$$\lim_n (a_n b_n) = \lim_n a_n \cdot \lim_n b_n.$$

16.25 Proposition. *Let (a_n) be a sequence converging to a. Then $(|a_n|)$ converges to $|a|$.*

Proof. This follows from $||a_n| - |a|| \leq |a_n - a|$. □

We have used: $|x - y| \geq ||x| - |y||$. This is a consequence of the triangle inequality: $|x| \leq |x - y| + |y|$ and so $|x| - |y| \leq |x - y|$, and similarly $|y| - |x| \leq |x - y|$.

The rule is

$$\lim_n |a_n| = |\lim_n a_n|.$$

16.26 Proposition. *Let (a_n) be a sequence converging to a and suppose that $a \neq 0$ and $a_n \neq 0$ for all n. Then the sequence $(\frac{1}{a_n})$ converges to $\frac{1}{a}$.*

Proof. We prove that $(\frac{1}{a_n} - \frac{1}{a})$ is a null sequence. We have

$$\frac{1}{a_n} - \frac{1}{a} = \frac{1}{a a_n}(a - a_n).$$

The sequence $(\frac{1}{a}(a - a_n))$ is a null sequence. Proposition 16.19 can be applied as soon as we have proved the sequence $(\frac{1}{a_n})$ to be bounded. There is an $N \in \mathbb{N}$ with $|a_n - a| < \frac{|a|}{2}$ for all $n \geq N$. Then $|a| \leq |a_n| + |a - a_n| \leq |a_n| + \frac{|a|}{2}$ and so $|a_n| \geq \frac{|a|}{2}$ for all $n \geq N$. For this n we have $\frac{1}{|a_n|} \leq \frac{2}{|a|}$. So the sequence $(\frac{1}{a_n})$ is bounded. □

We have the rule

$$\lim_n \frac{1}{a_n} = \frac{1}{\lim_n a_n}.$$

Everything in this formula has to be defined.

From lemma 16.9 follows directly:

16.27 Lemma. *Subsequences of a convergent sequence are convergent.* □

If a subsequence converges, then the sequence itself does not necessarily converge. It does if the sequence is ascending (or descending). That will be proposition 16.30.

16.28 Definition. A sequence (a_n) of rational numbers is called *ascending* (respectively *descending*) if $a_{n+1} \geq a_n$ (respectively $a_{n+1} \leq a_n$) for all indices n.

16.29 Lemma. *Let (a_n) be an ascending convergent sequence of rational numbers. Then $a_n \leq \lim_n a_n$ for all n.*

Proof. Let $a = \lim_n a_n$. Suppose there is an m with $a_m > a$. Then for all $n \geq m$ we have $a_n \geq a_m > a$, that is $a_n - a \geq a_m - a$. It would follow that for $\varepsilon = a_m - a$ there is no N such that $|a_n - a| < \varepsilon$ for all $n \geq N$. □

16.30 Proposition. *Let (a_n) be an ascending sequence of rational numbers with a convergent subsequence $(a_{i(n)})$. Then (a_n) converges as well and its limit is equal to the limit of this subsequence.*

Proof. Let a be the limit of the sequence $(a_{i(n)})$. Let $\varepsilon > 0$. Then there exists, see also lemma 16.29, an M with $0 \leq a - a_{i(n)} < \varepsilon$ for all $n \geq M$. Take $N = i(M)$. Let $n \geq N$. There is an $M' > M$ with $i(M') > n$ and so

$$a_{i(M)} \leq a_n \leq a_{i(M')} \leq a,$$

that is

$$0 \leq a - a_{i(M')} \leq a - a_n \leq a - a_{i(M)} < \varepsilon.$$

So (a_n) converges to a. □

Series

16.31 Terminology and notation. The partial sum sequence (s_n) of a sequence (a_n) is the sequence given by

$$s_n = \sum_{k=0}^{n-1} a_k.$$

The sequence (a_n) is the difference sequence of (s_n):

$$a_n = s_{n+1} - s_n.$$

The sequence s_n is often referred to as a *series*. The a_n are the *terms* of the series. If (s_n) converges, then for the limit the following notation is used:

$$\sum_{n=0}^{\infty} a_n = \lim_n s_n.$$

The terms in the sequence (a_n) are indexed by $\mathbb{N}$ and for s_n the indexing is chosen in such a way that s_n is the sum of the first n terms. Other choices for the relation between the indices of the sequence and the sequence of its partial sums are possible and in many cases another choice might even be preferable. It is customary to use the notation

$$\sum_{n=m}^{\infty} a_n$$

for both the series and the limit of the series.

16.32 Proposition. *Suppose the series (s_n) with terms a_n converges. Then the sequence (a_n) is a null sequence.*

Proof. This is a reformulation of proposition 16.22: (a_n) is the difference sequence of its partial sum sequence. □

16.33 Definition. If the sequence (a_n) is a geometric sequence, then its partial sum sequence (s_n) is called a *geometric series*.

A geometric series is given by its first term a and a ratio r. So the terms of the series are ar^n. If the geometric series (s_n) with $s_n = \sum_{k=0}^{n-1} ar^k$ converges, then by proposition 16.32 the sequence (ar^n) is a null sequence and so $|r| < 1$ (if $a \neq 0$). For geometric series the converse of proposition 16.32 does hold:

16.34 Proposition. *Let a and r be rational numbers with $|r| < 1$. Then the geometric series with first term a and ratio r converges to*

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}.$$

Proof. By theorem 8.5 we have $s_n = \frac{a(1-r^n)}{1-r}$. The sequence (r^n) is a null sequence and so the sequence (s_n) converges. The limit is

$$\sum_{n=0}^{\infty} ar^n = \lim_n \frac{a(1-r^n)}{1-r} = \frac{a}{1-r}.$$ □

16.4 Base g Expansions

As noticed in section 16.1 a rational number r can be approximated by decimal fractions: the sequence (a_n) with $a_n = \frac{\lfloor 10^n r \rfloor}{10^n}$ converges to r:

$$0 \le r - \frac{\lfloor 10^n r \rfloor}{10^n} = \frac{10^n r - \lfloor 10^n r \rfloor}{10^n} < \frac{1}{10^n}.$$

In the decimal notation every next term improves the approximation of r: an extra digit is added. We will now focus on that extra digit, so on the difference sequence of (a_n). We will do this right away for a g-adic notation instead of just the decimal one.

We fix a base $g \ge 2$. Let r be a rational number with $0 \le r < 1$. Then the sequence (a_n) with $a_n = \frac{\lfloor g^n r \rfloor}{g^n}$ converges to r:

$$0 \le r - \frac{\lfloor g^n r \rfloor}{g^n} = \frac{g^n r - \lfloor g^n r \rfloor}{g^n} < \frac{1}{g^n}.$$

Consider the sequence (b_n) with $b_n = g^n r - \lfloor g^n r \rfloor$. We have $0 \le b_n < 1$ for all $n \in \mathbb{N}$ and $r - a_n = \frac{b_n}{g^n}$.

For all $n \in \mathbb{N}$:

$$\begin{aligned} g b_n - \lfloor g b_n \rfloor &= g^{n+1} r - g\lfloor g^n r \rfloor - \lfloor g^{n+1} r - g\lfloor g^n r \rfloor \rfloor = g^{n+1} r - \lfloor g^{n+1} r \rfloor \\ &= g^{n+1} - g\lfloor g^n r \rfloor - \lfloor g^{n+1} r \rfloor + g\lfloor g^n r \rfloor \rfloor = b_{n+1}. \end{aligned}$$

So the sequence (b_n) is the course of $b_0 = r - \lfloor r \rfloor = r$ under the transformation

$$\gamma \colon \mathbb{Q} \to \mathbb{Q}, \quad x \mapsto gx - \lfloor gx \rfloor,$$

that is $b_n = \gamma^n(r)$. From $\gamma(r) = gr - \lfloor gr \rfloor$ follows $r = \frac{\lfloor gr \rfloor}{g} + \frac{\gamma(r)}{g}$. We now have

$$\begin{aligned} r &= \frac{\lfloor gr \rfloor}{g} + \frac{\gamma(r)}{g} = \frac{\lfloor gr \rfloor}{g} + \frac{\lfloor g\gamma(r) \rfloor}{g^2} + \frac{\gamma^2(r)}{g^2} \\ &= \frac{\lfloor gr \rfloor}{g} + \frac{\lfloor g\gamma(r) \rfloor}{g^2} + \frac{\lfloor g\gamma^2(r) \rfloor}{g^3} + \frac{\gamma^3(r)}{g^3} \\ &\vdots \\ &= \Big(\sum_{k=1}^{n} \frac{\lfloor g\gamma^{k-1}(r) \rfloor}{g^k} \Big) + \frac{\gamma^n(r)}{g^n} \end{aligned}$$

and so

$$\lfloor g^n r \rfloor = \Big\lfloor \sum_{k=1}^{n} \lfloor g\gamma^{k-1}(r) \rfloor g^{n-k} + \gamma^n(r) \Big\rfloor = \sum_{k=1}^{n} \lfloor \gamma^{k-1}(r) \rfloor g^{n-k}.$$

It follows that

$$r = \lim_n a_n = \lim_n \frac{\lfloor g^n \rfloor}{g^n} = \lim_n \sum_{k=1}^{n} \frac{\lfloor g\gamma^{k-1}(r) \rfloor}{g^k} = \sum_{k=1}^{\infty} \frac{\lfloor g\gamma^{k-1}(r) \rfloor}{g^k}.$$

16.35 Definition. Let r be a rational number with $0 \leq r < 1$ and let γ be the transformation of $\mathbb{Q}$ defined above. Then the sequence $(\lfloor g\gamma^{n-1}(r) \rfloor)_{n \geq 1}$ is called the *g-adic expansion* of r.

In other words: if (c_n) is the g-adic expansion of r, then

$$r = \sum_{n=1}^{\infty} \frac{c_n}{g^n}.$$

If the base g is equal to ten, then we usually write $r = 0.c_1c_2c_3\ldots$.

16.36 Example. Computation of the decimal expansion of $\frac{1}{7}$:

$$\begin{aligned}
\tfrac{10}{7} &= 1 + \tfrac{3}{7}, \\
\tfrac{30}{7} &= 4 + \tfrac{2}{7}, \\
\tfrac{20}{7} &= 2 + \tfrac{6}{7}, \\
\tfrac{60}{7} &= 8 + \tfrac{4}{7}, \\
\tfrac{40}{7} &= 5 + \tfrac{5}{7}, \\
\tfrac{50}{7} &= 7 + \tfrac{1}{7}.
\end{aligned}$$

The course of $\frac{1}{7}$ under γ is the repeating sequence $(\overline{\frac{1}{7}, \frac{3}{7}, \frac{2}{7}, \frac{6}{7}, \frac{4}{7}, \frac{5}{7}})$. The decimal expansion is the repeating sequence $(\overline{1, 4, 2, 8, 5, 7})$ and so $\frac{1}{7} = 0.\overline{142857}$.

Let's also compute the binary expansion of $\frac{1}{7}$:

$$\begin{aligned}
\tfrac{2}{7} &= 0 + \tfrac{2}{7}, \\
\tfrac{4}{7} &= 0 + \tfrac{4}{7}, \\
\tfrac{8}{7} &= 1 + \tfrac{1}{7}.
\end{aligned}$$

So the binary notation for $\frac{1}{7}$ is $0.\overline{001}$.

Obviously the g-adic expansion of a rational number repeats.

16.37 Proposition. *Let r be a rational number with $0 \leq r < 1$. Then the g-adic expansion of r repeats. The length of the least period is at most equal to the least denominator of r.*

Proof. Write $r = \frac{a}{b}$ with $a \in \mathbb{Z}$, $b \in \mathbb{N}^+$ and $\gcd(a, b) = 1$. Then

$$\gamma\Big(\frac{a}{b}\Big) = \frac{ga}{b} - \Big\lfloor \frac{ga}{b} \Big\rfloor = \frac{r_b(ga)}{b},$$

where $r_b(ga)$ is the remainder of ga after division by b. So the course of r under γ is a sequence in $\{\frac{0}{b}, \frac{1}{b}, \dots, \frac{b-1}{b}\}$, a set consisting of b elements. It follows that the course repeats with a period of length $\le b$. The g-adic expansion $(\lfloor g\gamma^{n-1}(r)\rfloor)$ repeats with a period of the same length. □

From the proof it is clear that more can be said about the g-adic expansion of rational numbers.

16.38 Proposition. *Let $r = \frac{a}{b}$ with $a \in \mathbb{Z}$, $b \in \mathbb{N}^+$, $a < b$, $\gcd(a, b) = 1$ and $\gcd(g, b) = 1$. Then the g-adic expansion of r is purely repeating with a least period of length $\mathrm{o}_b(g)$, the least $k \in \mathbb{N}^+$ such that $g^k \equiv 1 \pmod b$ (see definition 13.20).*

Proof. Under the transformation γ a fraction $\frac{k}{b}$ is mapped to $\frac{r_b(gk)}{b}$. For the numerators this corresponds to the transformation $\mathbb{N}_b \to \mathbb{N}_b$, $k \mapsto r_b(gk)$ and also with the multiplication by $\overline{g}$:

$$\sigma_g \colon \mathbb{Z}/b \to \mathbb{Z}/b,\ \overline{k} \mapsto \overline{g} \cdot \overline{k}.$$

Since $\gcd(a, b) = 1$ we have $\overline{a} \in \mathbb{Z}/b^*$ and since $\gcd(g, b) = 1$ the transformation σ_g is a permutation. The length of the orbit of $\overline{a}$ under σ_g equals the order of g modulo b. □

> Since $\mathrm{o}_b(g)$ divides $\varphi(b)$, the length of the period is a divisor of the totient of b for every base g relatively prime to b.

16.39 Example. See example 16.36. We computed the decimal and the binary expansion of the rational number $\frac{1}{7}$. The order of 10 modulo 7 equals 6, and modulo 2 it equals 3. In both cases the length of the period is a divisor of $\varphi(7) = 6$.

So rational numbers have a repeating g-adic expansion. Conversely, if $(c_n)_{n\ge1}$ is a repeating sequence in $\mathbb{N}_g$, then it represents a rational number:

16.40 Proposition. *Let $g \in \mathbb{N}$ with $g \ge 2$. Let $(c_n)_{n\ge1}$ be a repeating sequence in $\mathbb{N}_g$. Then the sequence $(a_n)_{n\ge1}$ with $a_n = \sum_{k=1}^{n} \frac{c_k}{g^k}$ converges in $\mathbb{Q}$.*

Proof. There are $m, N \in \mathbb{N}^+$ such that $c_{n+m} = c_n$ for all $n \ge N$, that is

$$(c_n)_{n\ge1} = (c_1, \dots, c_{N-1}, \overline{c_N, \dots, c_{N+m-1}}).$$

First we show that the subsequence (d_n) with $d_n = a_{N-1+nm}$ converges. The sequence $(d_n - d_0)$ is a geometric series with ratio $\frac{1}{g^m}$:

$$(d_{n+1} - d_0) - (d_n - d_0) = d_{n+1} - d_n = a_{N-1+(n+1)m} - a_{N-1+nm}$$

$$= \sum_{k=0}^{m-1} \frac{c_{N+nm+k}}{g^{N+nm+k}} = \sum_{k=0}^{m-1} \frac{c_{N+k}}{g^{N+nm+k}} = \frac{1}{g^{nm}} \sum_{k=0}^{m-1} \frac{c_{N+k}}{g^{N+k}}.$$

By proposition 16.34 the sequence $(d_n - d_0)$ converges and so does (d_n). Since the ascending sequence (a_n) has a convergent subsequence, it converges by proposition 16.30. □

16.41 Example. For the repeating sequence $(c_n)_{n\geq 1} = (2, 7, \overline{5, 0, 7})$ in $\mathbb{N}_{10}$ we compute the fraction represented by $\sum_{n=1}^{\infty} \frac{c_n}{10^n} = 0.27\overline{507}$. First we compute $0.00\overline{507}$. This is the sum of the geometric series with first term 0.00507 and ratio 0.001. So $0.00\overline{507} = \frac{0.00507}{0{,}999} = \frac{507}{99900} = \frac{169}{33300}$. And so $0.27\overline{507} = 0.27 + 0.00\overline{507} = \frac{27}{100} + \frac{169}{33300} = \frac{458}{1665}$.

A sequence $(c_n)_{n\geq 1}$ in $\mathbb{N}_g$ with a $g-1$-tail repeats and so $\sum_{n=1}^{\infty} \frac{c_n}{g^n}$ exists:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{c_n}{g^n} &= \frac{c_1}{g} + \frac{c_2}{g^2} + \cdots + \frac{c_{N-1}}{g^{N-1}} + \sum_{n=N}^{\infty} \frac{g-1}{g^n} \\
&= \frac{c_1}{g} + \frac{c_2}{g^2} + \cdots + \frac{c_{N-1}}{g^{N-1}} + \frac{\frac{g-1}{g^N}}{1-\frac{1}{g}} = \frac{c_1}{g} + \frac{c_2}{g^2} + \cdots + \frac{c_{N-1}}{g^{N-1}} + \frac{1}{g^{N-1}} \\
&= \frac{c_1}{g} + \frac{c_2}{g^2} + \cdots + \frac{c_{N-1}+1}{g^{N-1}}.
\end{aligned}$$

So the g-adic expansion of this number is $(c_1, c_2, \ldots, c_{N-1}+1, 0, 0, 0, \ldots)$. A $g-1$-tail will not occur in a g-adic expansion of a rational number.

If $(c_n)_{n\geq 1}$ is a sequence in $\mathbb{N}_g$ with not all c_n equal to $g-1$ for which $(\sum_{k=1} \frac{c_k}{g^k})$ converges, say $r = \sum_{n=1}^{\infty} \frac{c_n}{g^n}$, then $\gamma(r) = (c_{n+1})_{n\geq 1}$ and $c_1 = \lfloor \gamma(r) \rfloor$. If the sequence (c_n) has no $g-1$-tail, then (c_n) is the g-adic expansion of r. In particular the sequence (c_n) repeats.

Thus we have a correspondence between repeating sequences without a $g-1$-tail in $\mathbb{N}_g$ and rational numbers r with $0 \leq r < 1$.

If (c_n) is a nonrepeating sequence in $\mathbb{N}_{10}$, for example the sequence

$$(0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, \ldots),$$

then the sequence (a_n) with $a_n = \sum_{k=1}^{n} \frac{c_n}{10^n}$ is a sequence which does not converge in $\mathbb{Q}$. In the next chapter we will extend $\mathbb{Q}$ to the field $\mathbb{R}$ of the real numbers. In that field this sequence does have a limit.

Python

The function `g_repr(a, b, g)` returns the `g`-adic expansion of the fraction represented by `(a, b)`, where $a, b \in \mathbb{N}^+$ and $a < b$. The expansion is a list of two lists: the first one is the initial part and the second the shortest period.

arithmetics.py

```
def g_expand(a, b, g):
    nrs = []
    exp = []
    while a not in nrs:
        nrs.append(a)
        (c, a) = divmod(g * a, b)
        exp.append(c)
    i = nrs.index(a)
```

```
>>> g_expand(1, 7, 10)
[[], [1, 4, 2, 8, 5, 7]]
>>> g_expand(1, 7, 2)
[[], [0, 0, 1]]
>>> g_expand(121, 725, 5)
[[0, 4], [0, 4, 1, 2, 3, 3, 4, 4, 0, 3, 2, 1, 1, 0]]
```

Given a repeating sequence, the function `rat(nrlist1, nrlist2, g)` computes the rational number it represents, see example 16.41.

arithmetics.py

```
from functools import reduce

def nat(nrlist, g):
    if nrlist == []: return 0
    def sumg(a ,b): return (a * g) + b
    return reduce(sumg, nrlist)

def rat(nrlist1, nrlist2, g):
    a, b = nat(nrlist1, g), nat(nrlist2, g)
    k, l = len(nrlist1), len(nrlist2)
    return simplify(a * (g**l - 1) + b, (g**l - 1) * g**k)
```

```
>>> rat([], [1, 4, 2, 8, 5, 7], 10)
(1, 7)
>>> rat([2], [0], 10)
(1, 5)
>>> rat([2, 7], [5, 0, 7], 10)
(458, 1665)
>>> rat([], [0, 0, 1], 2)
(1, 7)
```

16.5 Cauchy Sequences

We want more sequences in $\mathbb{Q}$ to converge, that is, we want to extend $\mathbb{Q}$ with new limits. But what is the right criterion for having a limit in the extended number system? If a sequence converges, its difference sequence is a null sequence. However, there are sequences with a null sequence as difference sequence, which do diverge whatever the extension will be. An example of this phenomenon:

16.42 Example. The sequence (h_n) with $h_n = \sum_{k=1}^{n} \frac{1}{k}$ is called the *harmonic series*. The number h_n is the sum of the first n terms of the sequence $(\frac{1}{n})_{n\geq 1}$. This last sequence is a null sequence. We will show that the harmonic sequence diverges.

Consider the subsequence $(d_n)_{n\geq 0}$ with $d_n = h_{2^n}$. We have $d_0 = h_1 = 1$ and

$$d_{n+1} = d_n + \sum_{k=2^n+1}^{2^{n+1}} \frac{1}{k} \geq d_n + \sum_{k=2^n+1}^{2^{n+1}} \frac{1}{2^{n+1}} = d_n + \tfrac{1}{2}.$$

By induction it follows that $d_n \geq 1 + \frac{n}{2}$. So (d_n) diverges and so does (h_n).

This section is about sequences in $\mathbb{Q}$ which should converge: the Cauchy sequences.

16.43 Definition. A sequence (a_n) of rational numbers is called a *Cauchy sequence* if for each $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $|a_m - a_n| < \varepsilon$ for all $m, n \geq N$.

Indeed, converging sequences are Cauchy sequences:

16.44 Proposition. *Let (a_n) be a convergent sequence in $\mathbb{Q}$. Then (a_n) is a Cauchy sequence.*

Proof. Let a be the limit of (a_n) and let $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ such that $|a_n - a| < \frac{\varepsilon}{2}$ for all $n \geq N$. For $m, n \geq N$ we then have we $|a_n - a_m| \leq |a_n - a| + |a - a_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. So (a_n) is a Cauchy sequence. □

> In the definition of Cauchy sequence the notion of limit does not occur. In the next chapter we will extend $\mathbb{Q}$ to the field $\mathbb{R}$. If a sequence in $\mathbb{Q}$ converges in $\mathbb{R}$, then then it is a Cauchy sequence in $\mathbb{R}$ and so it is in $\mathbb{Q}$. We will see that in fact in $\mathbb{R}$ all Cauchy sequences converge.

Also Cauchy sequences behave well under the operations of addition, multiplication and inversion.

16.45 Proposition. *Let (a_n) and (b_n) be Cauchy sequences. Then so is $(a_n + b_n)$.*

Proof. Let ε be any positive number. Since (a_n) and (b_n) are Cauchy sequences, there is an $N \in \mathbb{N}$ such that both $|a_n - a_m| < \frac{\varepsilon}{2}$ and $|b_n - b_m| < \frac{\varepsilon}{2}$ for all $m, n \geq N$. Then

$$|a_n + b_n - a_m - b_m| \leq |a_n - a_m| + |b_n - b_m| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon.$$

So $(a_n + b_n)$ is a Cauchy sequence. □

16.46 Proposition. *Cauchy sequences are bounded.*

Proof. Let (a_n) be a Cauchy sequence. Then there is an $N \in \mathbb{N}$ such that $|a_n - a_m| < 1$ for all $m, n \geq N$. In particular $|a_n - a_N| < 1$ and so $|a_n| \leq |a_N| + |a_n - a_N| < |a_N| + 1$ for all $n \geq N$. So is (a_n) bounded. □

16.47 Proposition. *Let (a_n) and (b_n) be Cauchy sequences. Then so is $(a_n b_n)$.*

Proof. We have

$$|a_n b_n - a_m b_m| = |a_n b_n - a_n b_m + a_n b_m - a_m b_m| \leq |a_n||b_n - b_m| + |a_n - a_m||b_m|.$$

Let ε be any positive number. Since (a_n) is a Cauchy sequence, it is bounded: there is a C such that $|a_n| \leq C$ for all n. Also the Cauchy sequence (b_n) is bounded: there is a D such that $|b_m| \leq D$ for all m. There is an $M \in \mathbb{N}$ such that $|a_n - a_m| < \frac{\varepsilon}{2D}$ for all $m, n \geq M$ and there is an $N \in \mathbb{N}$ such that $|b_n - b_m| < \frac{\varepsilon}{2C}$ for all $m, n \geq N$. We have for all $m, n \geq \max(M, N)$:

$$|a_n b_n - a_m b_m| < C \cdot \tfrac{\varepsilon}{2C} + D \cdot \tfrac{\varepsilon}{2D} = \varepsilon.$$ □

16.48 Lemma. *Let (a_n) be a Cauchy sequence which is not a null sequence. Then there is a $C > 0$ and an $N \in \mathbb{N}$ such that $|a_n| > C$ for all $n \geq N$.*

Proof. Because (a_n) is not a null sequence, there is an $\varepsilon > 0$ such that for every $N \in \mathbb{N}$ there is an $n \geq N$ with $|a_n| > \varepsilon$. Since (a_n) is a Cauchy sequence, there is an N with $|a_m - a_n| < \frac{\varepsilon}{2}$ for all $m, n \geq N$. It follows that there is an $M \geq N$ with $|a_M| > \varepsilon$. We have $|a_M| \leq |a_M - a_n| + |a_n|$ and so $|a_n| \geq |a_M| - |a_M - a_n| > \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}$ for all $n \geq M$. So take $C = \frac{\varepsilon}{2}$. □

16.49 Proposition. *Let (a_n) be a Cauchy sequence which is not a null sequence and let $a_n \neq 0$ for all n. Then $(\frac{1}{a_n})$ is a Cauchy sequence as well.*

Proof. By lemma 16.48 there exist a $C > 0$ and an $N \in \mathbb{N}$ such that $|a_n| > C$ for all $n \geq N$. Let ε be any positive number. Then there is a $K \in \mathbb{N}$ such that $|a_m - a_n| < C^2 \varepsilon$ for all $m, n \geq K$. For all $m, n \geq \max(K, N)$ we then have

$$\left| \frac{1}{a_n} - \frac{1}{a_m} \right| = \frac{1}{|a_n a_m|} \cdot |a_n - a_m| < \frac{1}{C^2} \cdot C^2 \varepsilon = \varepsilon.$$ □

The fact that addition, multiplication and inversion of Cauchy sequences again results in Cauchy sequences is of importance in the next chapter: consequences will be that the operations in $\mathbb{R}$ are easily defined and, moreover, the rules of arithmetic for $\mathbb{R}$ are easily verified.

We can strengthen lemma 16.48 further:

16.50 Proposition. *Let (a_n) be a Cauchy sequence which is not a null sequence. Then there is a $C > 0$ and an $N \in \mathbb{N}$ such that either $a_n > C$ for all $n \geq N$, or $a_n < -C$ for all $n \geq N$.*

Proof. By lemma 16.48 there is an $M \in \mathbb{N}$ and a $C > 0$ such that $|a_n| > C$ for all $n \geq M$. Since (a_n) is a Cauchy sequence, there is a $K \in \mathbb{N}$ such that $|a_n - a_m| < C$ for all $m, n \geq K$. Take $N = \max(M, K)$.

> Assume $a_N > C$. For all $n \geq N$ we have $a_N - a_n < C$ and so $a_n > a_N - C > 0$. It follows that $a_n = |a_n| > C$.
>
> Assume $a_N < -C$. For all $n \geq N$ we have $-C < a_N - a_n$ and so $a_n < a_N + C < 0$. It follows that $a_n = -|a_n| < -C$. □

The definition of Cauchy sequence might give the impression that it is often difficult to show that a sequence is a Cauchy sequence. In some situations however, it is easily done.

16.51 Theorem. *Let (a_n) be an ascending sequence of rational numbers and (b_n) a descending sequence of rational numbers. Let furthermore be given that $a_n \leq b_n$ for all $n \in \mathbb{N}$ and that the sequence $(b_n - a_n)$ is a null sequence. Then (a_n) and (b_n) are Cauchy sequences.*

Proof. Let $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ such that $0 \leq b_n - a_n < \varepsilon$ for all $n \geq N$. For $m, n \geq N$ with $m \leq n$ we then have

$$a_N \leq a_m \leq a_n \leq b_n \leq b_N$$

and so $0 \leq a_n - a_m < b_N - a_N < \varepsilon$. So (a_n) is a Cauchy sequence. Similarly, (b_n) is a Cauchy sequence. This also follows from $b_n = a_n + (b_n - a_n)$. □

16.52 Corollary. *Let $(c_n)_{n \geq 1}$ be a sequence in $\mathbb{N}_g$. Then $(a_n)_{n \geq 1}$ with $a_n = \sum_{k=1}^{n} \frac{c_k}{g^k}$ is a Cauchy sequence.*

Proof. The sequence (a_n) is ascending and the sequence $(a_n + \frac{1}{g^n})$ is descending:

$$a_{n+1} + \frac{1}{g^{n+1}} = a_n + \frac{c_{n+1}}{g^{n+1}} + \frac{1}{g^{n+1}} = a_n + \frac{c_{n+1} + 1}{g^{n+1}} \leq a_n + \frac{g}{g^{n+1}} = a_n + \frac{1}{g^n}.$$

From theorem 16.51 follows that (a_n) is a Cauchy sequence. □

> So a sequence like 0.1, 0.17, 0.172, 0.1720, 0.17207, ..., where an extra digit is added in each term, is a Cauchy sequence. If the sequence is repeating, then it converges to a rational number. If not, then it does not converge (in $\mathbb{Q}$), because the sequence would be the decimal expansion of its limit.

16.6 p-Adic Approximations

In this section for every prime number p an absolute value on $\mathbb{Q}$ is defined. These absolute values differ strongly from the ordinary absolute value we used so far.

16.53 Definition. Let p be a prime number and r a nonzero rational number. We define the *p-adic absolute value* $|r|_p$ of r as follows:

$$|r|_p = p^{-v_p(r)}.$$

And we define $|0|_p = 0$. Thus we have a map $\mathbb{Q} \to \mathbb{Q}^{\geq 0}, r \mapsto |r|_p$, the *p-adic absolute value on* $\mathbb{Q}$.

16.54 Proposition. *Let p be a prime number. For all $r, s \in \mathbb{Q}$ we have:*

(i) $|r|_p = 0 \iff r = 0$,
(ii) $|rs|_p = |r|_p \cdot |s|_p$,
(iii) $|r + s|_p \leq \max(|r|_p, |s|_p)$.
(iv) *If* $|r|_p \neq |s|_p$, *then* $|r + s|_p = \max(|r|_p, |s|_p)$.

Proof.

(i) This follows immediately from the definition.
(ii) This is a consequence of $v_p(rs) = v_p(r) + v_p(s)$ if $r, s \neq 0$. For $r = 0$ or $s = 0$ it is obvious.
(iii) Obvious for $r = 0$ or $s = 0$. We assume $r, s \neq 0$ and put $r = \frac{a}{c}$ and $s = \frac{b}{c}$ with $c \in \mathbb{N}^+$ and $a, b \in \mathbb{Z}$. Then (by part (ii)) to prove that $|a + b|_p \leq \max(|a|_p, |b|_p)$, that is, $v_p(a + b) \geq \min(v_p(a), v_p(b))$. We may assume that $v_p(a) \leq v_p(b)$. Then to prove $v_p(a + b) \geq v_p(a)$. Let $k = v_p(a)$. Then $p^k \mid a$. Also $p^k \mid b$, because $v_p(b) \geq v_p(a) = k$. So $p^k \mid a + b$, that is, $v_p(a + b) \geq k = v_p(a)$.
(iv) We use the notation of the proof of part (iii). Now to prove the equality $v_p(a + b) = \min(v_p(a), v_p(b))$. We assume that $v_p(a) < v_p(b)$. Then to prove $v_p(a + b) = v_p(a)$. Write $a = p^k a'$ and $b = p^k b'$. Then $v_p(a') = 0$ and $v_p(b') > 0$, that is, $p \nmid a'$ and $p \mid b'$. It follows that $v_p(a' + b') = 0$. So $v_p(a + b) = k + v_p(a' + b') = k = v_p(a)$. □

The third property is even stronger than what was required for absolute values. It follows that $r \mapsto |r|_p$ is an absolute value on $\mathbb{Q}$. The ordinary absolute value satisfies $|n| = n$ for all $n \in \mathbb{N}^+$. For the p-adic absolute value we have

$$|n|_p = |1 + \cdots + 1|_p \leq \max(|1|_p, \ldots, |1|_p) = 1.$$

An absolute value with this property is called *non-Archimedean*, and otherwise it is called *Archimedean*. The p-adic absolute value can be used to define the p-adic distance.

16.55 Definition. Let r and s be rational numbers. The *p-adic distance* $d_p(r,s)$ of r to s is the p-adic absolute value of the difference of r and s:

$$d_p(r,s) = |r-s|_p.$$

The p-adic distance is a metric on $\mathbb{Q}$, and by property (iii) even an *ultrametric*:

16.56 Proposition. *For all r, s and t in $\mathbb{Q}$:*

(i) $d_p(r,s) = 0 \iff r = s$,
(ii) $d_p(r,s) = d_p(s,r)$,
(iii) $d_p(r,t) \leq \max(d_p(r,s), d_p(s,t))$.

Proof. These rules follow directly from proposition 16.54. □

In the sections 16.2, 16.3 and 16.5 we introduced notions that are equally well applicable for any absolute value. Many of the properties we derived were consequences of the defining rules for absolute values only. For some of the properties the definition of the ordinary absolute value was used. The ordinary absolute value is closely connected to the ordering of $\mathbb{Q}$. The p-adic absolute value is of importance for the arithmetic of rational numbers. Again we will study limits of sequences, now with respect to the p-adic metric. For the proofs it often suffices to refer to proofs in the preceding sections.

16.6.1 p-adic convergence

16.57 Definition. A sequence (a_n) of rational numbers is called a *p-adic null sequence* if for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that

$$|a_n|_p < \varepsilon \text{ for all } n \geq N.$$

Put differently: (a_n) is a p-adic null sequence if and only if $(|a_n|_p)$ is a null sequence. (Compare with lemma 16.11. In this section by a null sequence, without 'p-adic', we mean a null sequence with respect to the ordinary absolute value.)

16.58 Example. The sequence (p^n) is a p-adic null sequence, since $|p^n|_p = \frac{1}{p^n}$ and $(\frac{1}{p^n})$ is a null sequence. The sequence $(\frac{1}{n})$ is not a p-adic null sequence: $|\frac{1}{n}|_p = p^{v_p(n)}$ and in particular $|\frac{1}{p^n}|_p = p^n$

The numbers a for which the sequence (a^n) is a p-adic null sequence are easily determined.

16.59 Proposition. *Let $a \in \mathbb{Q}$. Then (a^n) is a p-adic null sequence if and only if $|a|_p < 1$.*

Proof. If $|a|_p \geq 1$, then $|a^n|_p = |a|_p^n \geq 1$ for all $n \in \mathbb{N}$. If $|a|_p < 1$, then the sequence $(|a|_p^n)$ is a null sequence by proposition 16.15. □

16.60 Proposition. *Let (a_n) and (b_n) be p-adic null sequences. Then so is (a_n+b_n).*

Proof. The sequences $(|a_n|_p)$ and $(|b_n|_p)$ are null sequences. So by proposition 16.16 the sequence $(|a_n|_p + |b_n|_p)$ is a null sequence and, since $|a_n + b_n|_p \leq |a_n|_p + |b_n|_p$, so is $(|a_n + b_n|_p)$. □

16.61 Definition. Let (a_n) be a sequence of rational numbers. We say that (a_n) *p-adically converges* to a rational number a if the sequence $(a_n - a)$ is a p-adic null sequence. Notation: $\lim_n^{(p)} a_n = a$.

For the p-adic absolute values of a p-adic convergent sequence the following is stronger then the equivalent of proposition 16.25:

16.62 Proposition. *Let (a_n) be a sequence of rational numbers converging p-adically to a. Then $(|a_n|_p)$ converges to $|a|_p$. If $a \neq 0$, then there is an $N \in \mathbb{N}$ such that $|a_n|_p = |a|_p$ for all $n \geq N$.*

Proof. For $a = 0$ the first part of the proposition is just the definition of p-adic null sequence. For $a \neq 0$ the first part follows from the second. We prove the second part. Because (a_n) converges p-adically there is an $N \in \mathbb{N}$ with $|a_n - a|_p < |a|_p$ for all $n \geq N$. From proposition 16.54(iv) it then follows that for this n we have $|a_n|_p = |a + (a_n - a)|_p = |a|_p$. □

> So: if a sequence converges p-adically and it is not a p-adic null sequence, then the p-adic absolute values of the terms eventually are equal.

16.63 Proposition. *Let (a_n) be a sequence converging p-adically to a and (b_n) be a sequence converging p-adically to b. Then the sequence $(a_n + b_n)$ converges p-adically to $a + b$.*

Proof. The sequence $(a_n + b_n - a - b)$ is the sum of two p-adic null sequences. □

16.64 Proposition. *Let (a_n) be a sequence converging p-adically to a and (b_n) a sequence converging p-adically to b. Then the sequence $(a_n b_n)$ converges p-adically to ab.*

Proof. The sequence $(a_n b_n - ab)$ is the sum of the sequences $(a_n(b_n - b))$ and $((a_n - a)b)$ and these both are p-adic null sequences. □

16.65 Proposition. *Let (a_n) be a sequence converging p-adically to a and let $a \neq 0$ and also $a_n \neq 0$ for all n. Then the sequence $(\frac{1}{a_n})$ converges p-adically to $\frac{1}{a}$.*

Proof. To prove that $(|\frac{1}{a_n} - \frac{1}{a}|_p)$ is a null sequence. We have:

$$\left|\frac{1}{a_n} - \frac{1}{a}\right|_p = \frac{1}{|a|_p |a_n|_p} |a - a_n|_p.$$

The sequence $(\frac{1}{|a|_p}(|a - a_n)|_p)$ is a null sequence. By proposition 16.62 the terms of the sequence $(\frac{1}{|a_n|_p})$ are equal to $\frac{1}{|a|_p}$ for large n. □

16.66 Notation. If the partial sum sequence (s_n) of a sequence (a_n) converges p-adically, then we use for the p-adic limit the following notation:

$$\sum_{n=0}^{\infty}{}^{(p)}\, a_n = \lim_n{}^{(p)} s_n.$$

Also in the p-adic case a geometric series converges if the absolute value of the ratio is less than 1. The proof is analogous.

16.67 Proposition. *Let a and r be rational numbers $|r|_p < 1$. Then the partial sum sequence of the geometric series with first term a and ratio r converges p-adically and*

$$\sum_{n=0}^{\infty}{}^{(p)}\, ar^n = \frac{a}{1-r}. \qquad \square$$

16.68 Example. Since $|p|_p = \frac{1}{p} < 1$, we have

$$\sum_{n=0}^{\infty}{}^{(p)}\, p^n = \frac{1}{1-p}.$$

16.6.2 p-Adic expansions

Rational numbers have repeating base g expansions for any base g. Thus the rational number is obtained as a limit of rational numbers having only powers of the base g in the denominator. This limit is with respect to the ordinary absolute value. In the p-adic case it is natural to take p as a base. Natural numbers have a p-adic notation. For rational numbers r with $v_p(r) \geq 0$ we have p-adic expansions. However these expansions are not to the right, but to the left. Division by p^n means for the p-adic notation that the point moves n places to the left.

16.69 Definition. A rational number r is called *p-adically integral* if $|r|_p \leq 1$. The set of p-adic integers we denote by $\mathbb{Z}_{(p)}$.

> Thus the set $\mathbb{Z}_{(p)}$ is a subset of $\mathbb{Q}$ containing $\mathbb{Z}$. The elements of $\mathbb{Z}_{(p)}$ can be written as fractions of integers, the denominator not being a multiple of p. This set is closed under addition and multiplication, that is, if $r, s \in \mathbb{Z}_{(p)}$, then also $r + s, rs \in \mathbb{Z}_{(p)}$. It is easily verified that the set $\mathbb{Z}_{(p)}$ with these operations is an integral domain. The subset $\mathbb{Z}_{(p)}^*$ of invertible elements consists of the rational numbers r satisfying $|r|_p = 1$. These numbers can be written as a fraction of integers both not being multiples of p.

16.70 Proposition. *Let $r \in \mathbb{Z}_{(p)}$. Then there is a unique $c \in \mathbb{N}_p$ such that $|r - c|_p < 1$.*

Proof. Write $r = \frac{a}{b}$ with $p \nmid b$. Since $\gcd(b,p) = 1$, there are $x, y \in \mathbb{Z}$ such that $a = xb + yp$. There is a unique pair x, y with $x \in \mathbb{N}_p$ satisfying this equation. We now have $\frac{a}{b} = x + \frac{y}{b}p$. Take $c = x$. Then $r - c = \frac{y}{b}p$ and so $|r-c|_p = |\frac{y}{b}|_p \cdot |p|_p \leq \frac{1}{p} < 1$. □

16.71 Definition. Let $r \in \mathbb{Z}_{(p)}$. The unique $c \in \mathbb{N}_p$ with $|r-c|_p < 1$ is called the *remainder* of r after division by p. Notation: $c = [r]_p$.

We now have a transformation of $\mathbb{Z}_{(p)}$:

$$\gamma_p \colon \mathbb{Z}_{(p)} \to \mathbb{Z}_{(p)},\ r \mapsto \frac{r - [r]_p}{p}.$$

16.72 Definition. Let $r \in \mathbb{Z}_{(p)}$. The sequence $(c_n)_{n\geq 0}$ with $c_n = [\gamma_p^n(r)]_p$ in $\mathbb{N}_p$ is called the *p-adic expansion* of r.

16.73 Proposition. *Let (c_n) be the p-adic expansion of $r \in \mathbb{Z}_{(p)}$. Then*

$$r = \lim_n {}^{(p)} \sum_{k=0}^{n} c_k p^k = \sum_{k=0}^{\infty}{}^{(p)}\, c_k p^k.$$

Proof. We have

$$\begin{aligned} r &= c_0 + \gamma_p(r)p = c_0 + c_1 p + \gamma_p^2(r)p^2 = \cdots \\ &= c_0 + c_1 p + c_2 p^2 + \cdots + c_n p^n + \gamma_p^{n+1}(r)p^{n+1}. \end{aligned}$$

So $|r - \sum_{k=0}^{n} c_k p^k|_p = |\gamma_p^{n+1} p^{n+1}|_p \leq \frac{1}{p^{n+1}} < \frac{1}{p^n}$. □

The p-adic expansion of a p-adically integral rational number r repeats. We will prove this for r with $-1 < r \leq 0$. In that case the repetition is pure. The general case then easily follows.

16.74 Proposition. *Let $r \in \mathbb{Z}/(p)$ with $-1 < r \leq 0$. Then the p-adic expansion of r repeats purely.*

Proof. Write $r = -\frac{a}{b}$ with $p \nmid b$, $a \in \mathbb{N}_b$ and $\gcd(a,b) = 1$. Then $b\gamma_p(r) = \frac{-a - b[r]_p}{p} \in \mathbb{Z}$. From $0 \leq [r]_p \leq p-1$ follows $0 \leq b[r]_p \leq bp - b$ and so also $a \leq a + b[r]_p \leq bp - b + a < bp$. So $\gamma_p(r) = -\frac{a'}{b}$ with $a' = \frac{a + b[r]_p}{p} \in \mathbb{Z}$ and $a' \in \mathbb{N}_p$. Furthermore, $pa' \equiv a \pmod{b}$, so in $\mathbb{Z}/b^*$ we have $\overline{a'} = \overline{p}^{-1}\overline{a}$. So the expansion repeats purely with a period of length $\mathrm{o}_b(p)$. □

16.75 Example. We compute the 5-adic expansion of $\frac{1}{7}$. First we compute the 5-adic expansion of $\frac{1}{7} - 1 = -\frac{6}{7}$. Following the proof of proposition 16.74, we get consecutively:

$$-\tfrac{6}{7} = 2 - \tfrac{4}{7} \cdot 5$$

$$
\begin{aligned}
-\tfrac{4}{7} &= 3 - \tfrac{5}{7}\cdot 5\\
-\tfrac{5}{7} &= 0 - \tfrac{1}{7}\cdot 5\\
-\tfrac{1}{7} &= 2 - \tfrac{3}{7}\cdot 5\\
-\tfrac{3}{7} &= 1 - \tfrac{2}{7}\cdot 5\\
-\tfrac{2}{7} &= 4 - \tfrac{6}{7}\cdot 5.
\end{aligned}
$$

Here in each line the second fraction is obtained from the first by multiplying by 3 (the inverse of 5 modulo 7) and taking the remainder after division of the result by 7. It follows that the 5-adic expansion of $-\frac{6}{7}$ is the sequence $(\overline{2,3,0,2,1,4})$. So the 5-adic expansion of $\frac{1}{7}$ is $(3,\overline{3,0,2,1,4,2})$.

The same computation as above, but a different notation:

$$
\begin{aligned}
\tfrac{20}{7} &= 2 + \tfrac{6}{7}\\
\tfrac{25}{7} &= 3 + \tfrac{4}{7}\\
\tfrac{5}{7} &= 0 + \tfrac{5}{7}\\
\tfrac{15}{7} &= 2 + \tfrac{1}{7}\\
\tfrac{10}{7} &= 1 + \tfrac{3}{7}\\
\tfrac{30}{7} &= 4 + \tfrac{2}{7}.
\end{aligned}
$$

From bottom to top this is exactly the computation used for the base 5 expansion of $\frac{6}{7}$: it is $0.\overline{412032}$. The bar indicates it is a period repeating to the right. The 5-adic expansion of $-\frac{6}{7}$ is $\overline{412032}$, where now the bar indicates a period repeating to the left.

> This is a special case of a general phenomenon for the p-adic expansion of $-\frac{a}{b}$, where $b \in \mathbb{N}^+$, $a \in \mathbb{N}_b$, $p \nmid b$ and $\gcd(a,b) = 1$: if $(\overline{c_1,\ldots,c_n})$ is the base p expansion of $\frac{a}{b}$, then $(\overline{c_n,\ldots,c_1})$ is the p-adic expansion of $-\frac{a}{b}$.

16.76 Proposition. *Let $(c_n)_{n\geq 0}$ be a repeating sequence in $\mathbb{N}_p$. Then the sequence (a_n) with $a_n = \sum_{k=0}^{n} c_k$ converges p-adically to a rational number.*

Proof. The convergence of the sequence a_n can be reduced to the convergence of a geometric series with ratio p^m, where m is the length of the period. □

Python

The function `p_adic(a, b, p)` returns the `p`-adic expansion of the rational number represented by `(a, b)`, where $a \in \mathbb{Z}$ and $b \in \mathbb{N}^+$ with $p \nmid b$.

arithmetics.py
```
def p_adic(a, b, p):
    u = modinv(b, p)
    nrs = []
    exp = []
    while a not in nrs:
        nrs.append(a)
        c = (a * u) % p
        a = (a - (c * b)) // p
        exp.append(c)
    i = nrs.index(a)
    return [exp[:i], exp[i:]]
```

```
>>> p_adic(-6, 7, 5)
[[], [2, 3, 0, 2, 1, 4]]
>>> p_adic(-131, 87, 17)
[[11], [2, 10, 8, 12]]
```

Given a repeating sequence, the function `prat(nrlist1, nrlist2, p)` computes the rational number it represents.

arithmetics.py
```
def ratp(nrlist1, nrlist2, p):
    nrlist1.reverse()
    nrlist2.reverse()
    a, b = nat(nrlist1, p), nat(nrlist2, p)
    k, l = len(nrlist1), len(nrlist2)
    return simplify(a * (p**l - 1) - (b * p**k), p**l - 1)
```

```
>>> ratp([11], [2,10, 8, 12], 17)
(-131, 87)
>>> ratp([], [2, 3, 0, 2, 1, 4], 5)
(-6, 7)
```

16.6.3 p-adic Cauchy sequences

We have the notion of Cauchy sequence with respect to the p-adic distance as well. In this case it can be simplified considerably.

16.77 Definition. A sequence (a_n) of rational numbers is called a *p-adic Cauchy sequence* if for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that $|a_n - a_m|_p \leq \varepsilon$ for all $m, n \geq N$.

16.78 Proposition. *A sequence in $\mathbb{Q}$ is a p-adic Cauchy sequence if and only if its difference sequence is a p-adic null sequence.*

Proof. If (a_n) is a p-adic Cauchy sequence, then clearly the difference sequence is a p-adic null sequence. Suppose conversely that $(a_{n+1} - a_n)$ is a p-adic null sequence. Let $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ such that $|a_{n+1} - a_n|_p < \varepsilon$ for all $n \geq N$. For $m, n \geq N$ with $m \leq n$ we then have

$$\begin{aligned} |a_n - a_m|_p &= |(a_n - a_{n-1}) + (a_{n-1} - a_{n-2}) + \cdots + (a_{m+1} - a_m)|_p \\ &\leq \max(|a_n - a_{n-1}|_p, |a_{n-1} - a_{n-2}|_p, \ldots, |a_{m+1} - a_m|_p) < \varepsilon. \end{aligned}$$

So (a_n) is a p-adic Cauchy sequence. □

16.79 Corollary. *Let $(c_n)_{n\geq 0}$ be a sequence in $\mathbb{N}_p$. Then the sequence $(a_n)_{n\geq 0}$ with $a_n = \sum_{k=0}^{n} c_n p^n$ is a Cauchy sequence.* □

p-Adically convergent sequences are p-adic Cauchy sequences. One might prove this as was done in the case of the ordinary absolute value (proposition 16.44), but since in this case the notion of Cauchy sequence is so simple, there is an easier proof.

16.80 Proposition. *p-Adically convergent sequences of rational numbers are p-adic Cauchy sequences.*

Proof. The difference sequence of a convergent sequence is a null sequence, so by proposition 16.78 it is a p-adic Cauchy sequence. □

The p-adic absolute values of the terms of a p-adic Cauchy sequences not being a p-adic null sequence are eventually constant. This generalizes proposition 16.62.

16.81 Lemma. *Let (a_n) be a p-adic Cauchy sequence which is not a p-adic null sequence. Then there is an $N \in \mathbb{N}$ such that $|a_n|_p = |a_{n+1}|_p$ for all $n \geq N$.*

Proof. Since (a_n) is not a p-adic null sequence, there is a $\varepsilon > 0$ such that for every $M \in \mathbb{N}$ there is a $N \geq M$ with $|a_N|_p > \varepsilon$. The sequence $(a_{n+1} - a_n)$ is a p-adic null sequence. So there is a $K \in \mathbb{N}$ such that $|a_{n+1} - a_n|_p < \varepsilon$ for all $n \geq K$. There is an $N \geq K$ with $|a_N|_p > \varepsilon$. For all $n \geq N$ we then have $|a_{n+1}|_p = |a_{n+1} - a_n + a_n|_p = |a_n|_p$. □

The fact that p-adic Cauchy sequences behave well under the operations addition, multiplication and inversion can be proved as was done in the ordinary case. Below shorter proofs are given based on proposition 16.78 and lemma 16.81. These properties will be used in chapter 18 where the field $\mathbb{Q}_p$ of the p-adic numbers is constructed.

16.82 Proposition. *Let (a_n) and (b_n) be p-adic Cauchy sequences. Then so is $(a_n + b_n)$.*

Proof. The sequence $(a_{n+1} + b_{n+1} - a_n - b_n)$ is the difference of two p-adic null sequences, $(a_{n+1} - a_n)$ and $(b_{n+1} - b_n)$, and so is a null sequence as well. □

16.83 Proposition. *Let (a_n) and (b_n) be p-adic Cauchy sequences. Then so is $(a_n b_n)$.*

Proof. We have

$$a_{n+1}b_{n+1} - a_n b_n = a_{n+1}(b_{n+1} - b_n) + (a_{n+1} - a_n)b_n.$$

By lemma 16.81 there is an $N \in \mathbb{N}$ such that $|a_{n+1}|_p = |a_n|_p$ for all $n \geq N$. Since $(b_{n+1} - b_n)$ is a p-adic null sequence, it follows that $(a_{n+1}(b_{n+1} - b_n))$ is a p-adic null sequence as well. Similarly, $((a_{n+1} - a_n)b_n)$ is a p-adic null sequence. So the sequence $(a_{n+1}b_{n+1} - a_n b_n)$ is a p-adic null sequence. □

16.84 Proposition. *Let (a_n) be a p-adic Cauchy sequence which is not a p-adic null sequence and let $a_n \neq 0$ for all n. Then also $(\frac{1}{a_n})$ is a p-adic Cauchy sequence.*

Proof. By lemma 16.81 there is an $N \in \mathbb{N}$ such that $|a_n|_p = |a_{n+1}|_p$ for all $n \geq N$. We have

$$\Big|\frac{1}{a_{n+1}} - \frac{1}{a_n}\Big|_p = \frac{1}{|a_n|_p |a_{n+1}|_p}|a_{n+1} - a_n|_p.$$

So for all $n \geq N$

$$\Big|\frac{1}{a_{n+1}} - \frac{1}{a_n}\Big|_p = \frac{1}{|a_N|_p^2}|a_{n+1} - a_n|_p.$$

So $(\frac{1}{a_{n+1}} - \frac{1}{a_n})$ is a p-adic null sequence. □

Exercises

1. (i) Show that the sequence (a_n) with $a_n = \frac{1}{2^n} - \frac{1}{5^n} + \frac{1}{7^n}$ is a null sequence.
   (ii) Determine an $N \in \mathbb{N}$ such that $|a_n| < \frac{1}{1000}$ for all $n \geq N$.

2. Prove that the sequence (a_n) with $a_n = \frac{(n!)^2}{(2n)!}$ is a null sequence.

3. (i) Prove that $2^n \geq n^2$ for all $n \in \mathbb{N}$ with $n \geq 4$.
   (ii) Prove that the sequence (a_n) with $a_n = \frac{n}{2^n}$ is a null sequence.

4. The sequence (a_n) with $a_n = \frac{2n^3+1}{3n^3+n+1}$ converges. Prove this and indicate the propositions used.

5. Determine $\lim_n \frac{2^n+3^n+4^n}{4^n-3^n-2^n}$.

6. Let $(c_n)_{n\geq 1}$ be a descending null sequence. The sequence $(a_n)_{n\geq 1}$ is defined by $a_n = \sum_{k=1}^{n}(-1)^{k+1}c_k$.
   (i) Let $N \in \mathbb{N}$. Prove that $a_{2N} \leq a_n \leq a_{2N+1}$ for all $n > 2N+1$.
   (ii) Prove that (a_n) is a Cauchy sequence.
   (iii) Prove that the sequence (a_n) defined by
   $$a_n = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \cdots + \tfrac{(-1)^{n+1}}{n} \quad \text{(for all } n \in \mathbb{N}\text{)}$$
   is a Cauchy sequence.
7. Determine the base g expansion of $\frac{1}{10}$ for $g = 2, 3, 4, 5, 6, 7$.
8. Write the rational number $8.0\overline{46}$ as a fraction of integers.
9. (i) Determine the base g expansion of the rational number $\frac{1}{2}$ for even bases g.
   (ii) Determine the base g expansion of the rational number $\frac{1}{2}$ for odd bases g.
10. Write the number with hexadecimal notation $2\mathrm{A}.2\overline{3\mathrm{F}}$ as a fraction of integers.
11. For a given g is $0.\overline{1}$ the g-adic notation of a rational number. Which number?
12. Let r be a rational number with $0 \leq r \leq 1$. How can the decimal notation of $1 - r$ be derived from the decimal notation of r?
13. There are 32 rational numbers with binary notation $0.\overline{c_1c_2c_3c_4c_5}$ and $c_1, c_2, c_3, c_4, c_5 \in \{0, 1\}$. Which numbers?
14. Show that the map γ described on page 335 restricts to a permutation of the set $\{\, r \in \mathbb{Q} \mid 0 \leq r < 1 \,\}$.
15. Determine the p-adic expansion of $\frac{1}{5}$ for $p = 2, 3, 7, 11, 13$.
16. The 2-adic notation of a rational number is $\overline{10110}1$. Write this number as a fraction of integers.
17. Determine $\lim_n^{(2)} 5^{2^n}$. (See exercise 16 of chapter 13.)
18. Let p be a prime number. Show that $\lim_n^{(p)} n! = 0$.
19. Let (a_n) be a Cauchy sequence having a subsequence converging in $\mathbb{Q}$. Prove that (a_n) converges in $\mathbb{Q}$. Verify that this holds in the p-adic situation as well.
20. Determine for all primes p the p-adic expansion of -1.
21. The 7-adic notation of a rational number is $\ldots\overline{304}12$. Write this number as a fraction of integers.
22. Let p be a prime number. Show that the map γ_p described on page 347 is a permutation of the set $\mathbb{Z}_{(p)}$.

17 The Real Numbers

A field with an absolute value can be extended to a field in which all Cauchy sequences converge, the so-called *completion* of the field. In this chapter we will do this for the field $\mathbb{Q}$ with the ordinary absolute value. The field thus obtained is the field $\mathbb{R}$ of the real numbers. The p-adic absolute values lead to other completions of $\mathbb{Q}$, for these see the next chapter. Because in $\mathbb{R}$ all Cauchy sequences converge we now have interesting examples of convergent sequences. In particular we consider infinite continued fractions. These are very well suited for approximation of real numbers by rationals. An important property of $\mathbb{R}$ is that the multiplicative group $\mathbb{R}^*$ strongly resembles the additive group of $\mathbb{R}$: we have in fact an isomorphism $\mathbb{R} \xrightarrow{\sim} \mathbb{R}^+$, where the addition of real numbers corresponds with the multiplication of positive real numbers.

17.1 The Construction of $\mathbb{R}$

We start with the field $\mathbb{Q}$ with its ordinary absolute value and want to construct an extension in which Cauchy sequences converge. The set $\mathbb{R}$ will be a set of classes of Cauchy sequences in $\mathbb{Q}$.

17.1.1 The set $\mathbb{R}$

In the field to be constructed two Cauchy sequences will have the same limit if and only if their difference is a null sequence. This leads to the following definition.

17.1 Definition. Cauchy sequences (a_n) and (b_n) in $\mathbb{Q}$ are called *equivalent* if the sequence $(a_n - b_n)$ is a null sequence. Notation: $(a_n) \sim (b_n)$. We denote the set of Cauchy sequences in $\mathbb{Q}$ by $\mathrm{CS}(\mathbb{Q})$. Thus the relation $\sim$ is a relation in the set $\mathrm{CS}(\mathbb{Q})$.

17.2 Proposition. *The relation $\sim$ in $\mathrm{CS}(\mathbb{Q})$ is an equivalence relation.*

Proof. The relation $\sim$ clearly is reflexive and symmetric. Transitivity follows easily from proposition 16.16. □

17.3 Definition. A *real number* is an equivalence class in $\mathrm{CS}(\mathbb{Q})$. Notation: the class of a Cauchy sequence (a_n) will be denoted by $[(a_n)]$. The set $\mathbb{R}$ is the set of the real numbers.

In this chapter often Greek letters are used to denote real numbers. For example $\alpha = [(a_n)]$: the real number α is represented by the Cauchy sequence (a_n) of rational numbers. There are many Cauchy sequences representing the same real number: other representatives are obtained by adding a null sequence to a given representative.

17.1.2 The field $\mathbb{R}$

Using the given construction of the set $\mathbb{R}$ addition and multiplication in $\mathbb{R}$ are easily defined and it is straightforward to prove that with these operations $\mathbb{R}$ is a field.

17.4 Definition. Let (a_n) and (b_n) be Cauchy sequences in $\mathbb{Q}$. The *sum* and the *product* of the real numbers $[(a_n)]$ and $[(b_n)]$ are defined by

$$[(a_n)] + [(b_n)] = [(a_n + b_n)]$$
$$[(a_n)] \cdot [(b_n)] = [(a_n b_n)].$$

If (a_n) and (b_n) are Cauchy sequences, then so are $(a_n + b_n)$ and $(a_n b_n)$, see proposition 16.45 and proposition 16.47. The definitions of sum and product should not depend on the choice of representatives. For example: if $(a_n) \sim (a'_n)$, then $(a_n b_n) \sim (a'_n b_n)$, that is $((a_n - a'_n)b_n)$ is a null sequence. This follows from proposition 16.19: $(a_n - a'_n)$ is a null sequence and (b_n) is bounded.

For $a \in \mathbb{Q}$ the constant sequence (a) is a Cauchy sequence. It represents a real number: $[(a)]$, the class of all sequences in $\mathbb{Q}$ converging to a. For rational numbers a and b we have:

$$\begin{aligned} [(a)] = [(b)] &\iff (a) \sim (b) \\ &\iff (a - b) \text{ is a null sequence} \\ &\iff a = b. \end{aligned}$$

So we have an injective map

$$\mathbb{Q} \to \mathbb{R},\ a \mapsto [(a)].$$

Addition and multiplication of $[(a)]$ and $[(b)]$ corresponds to addition and multiplication of the rational numbers a and b. So the numbers $[(a)]$ form a copy inside $\mathbb{R}$ of the field of rational numbers. That is why we will denote $[(a)]$ by a and consider $\mathbb{R}$ as an extension of $\mathbb{Q}$. In particular we have the elements 0 and 1 in $\mathbb{R}$. Moreover, we denote $[(-a_n)]$ by $-[(a_n)]$ (which is independent of the choice of the representative).

17.5 Theorem. *$\mathbb{R}$ together with the addition and the multiplication is a field.*

Proof. It is straightforward to show that $\mathbb{R}$ is a commutative ring. For example the distributivity: choose for real numbers α, β and γ representatives (a_n), (b_n) and (c_n), then $\alpha(\beta+\gamma) = [(a_n)]([(b_n)]+[(c_n)]) = [(a_n(b_n+c_n))]$ and also $\alpha\beta+\alpha\gamma = [(a_n)][(b_n)] + [(a_n)][(c_n)] = [(a_nb_n + a_nc_n)]$.

It only remains to prove that real numbers $\neq 0$ have inverses. Let $\alpha \in \mathbb{R}$ with $\alpha \neq 0$. Choose a representative (a_n) of α. The real number 0 is the class of the null sequences in $\mathbb{Q}$. So (a_n) is not a null sequence. By proposition 16.48 there is a $C > 0$ and a $N \in \mathbb{N}$ such that $|a_n| > C$ for all $n \geq N$. So we can assume that $a_n \neq 0$ (replace in the sequence terms 0 by 1, or take the sequence (a_{N+n})). From proposition 16.49 follows that the sequence $(\frac{1}{a_n})$ is a Cauchy sequence as well. We then have $[(a_n)][(\frac{1}{a_n})] = [(1)] = 1$. □

17.6 Definition. An $\alpha \in \mathbb{R}$ is called *irrational* if $\alpha \notin \mathbb{Q}$. (Here we consider $\mathbb{Q}$ as a part of $\mathbb{R}$.)

We will show that we have obtained many new numbers. The real numbers can be seen as limits of sequences of rational numbers. For this to be meaningful we have to extend the absolute value on $\mathbb{Q}$ to an absolute value on $\mathbb{R}$. First we extend the ordering of $\mathbb{Q}$ to $\mathbb{R}$.

17.1.3 The ordering of $\mathbb{R}$

We will use proposition 16.50. For a Cauchy sequence (a_n) not being a null sequence, there are two complementary possibilities:

a) There is a $C > 0$ and an $N \in \mathbb{N}$ such that $a_n > C$ for all $n \geq N$.
b) There is a $C > 0$ and an $N \in \mathbb{N}$ such that $a_n < -C$ for all $n \geq N$.

17.7 Definition. Let $\alpha \in \mathbb{R}$. Then α is called *positive* if $\alpha = [(a_n)]$ and for the Cauchy sequence (a_n) the first of the possibilities above holds for α. Otherwise, α is called *negative*.

> Again this does not depend on the choice of the representative. If there are for the Cauchy sequence (a_n) in $\mathbb{Q}$ a $C > 0$ and an $N \in \mathbb{N}$ with $a_n > C$ for all $n \geq N$ and (b_n) is a Cauchy sequence in $\mathbb{Q}$ with $[(a_n)] = [(b_n)]$, then $(a_n - b_n)$ is a null sequence and so there is an $M \in \mathbb{N}$ such that $|a_n - b_n| < \frac{C}{2}$. Then $b_n = a_n - (a_n - b_n) > C - \frac{C}{2} = \frac{C}{2}$ for all $n \geq \max(M, N)$.

17.8 Definition. Let α and β be real numbers. We define:

$$\alpha < \beta \iff \beta - \alpha \text{ is positive,}$$

and

$$\alpha \leq \beta \iff \alpha < \beta \text{ or } \alpha = \beta.$$

Instead of $\alpha < \beta$ we sometimes write $\beta > \alpha$. And $\beta \geq \alpha$ has the same meaning as $\alpha \leq \beta$.

The relation $\leq$ thus defined clearly is an extension of the relation $\leq$ on $\mathbb{Q}$ and is a ordering of the field $\mathbb{R}$ as well:

17.9 Proposition. *For all $\alpha, \beta, \gamma \in \mathbb{R}$:*

(i) $\alpha \leq \alpha$,
(ii) *if* $\alpha \leq \beta$ *and* $\beta \leq \alpha$, *then* $\alpha = \beta$,
(iii) *if* $\alpha \leq \beta$ *and* $\beta \leq \gamma$, *then* $\alpha \leq \gamma$,
(iv) *if* $\alpha \leq \beta$, *then* $\alpha + \gamma \leq \beta + \gamma$,
(v) *if* $\alpha \leq \beta$ *and* $\gamma > 0$, *then* $\alpha\gamma \leq \beta\gamma$.

Proof. Part (i) follows directly from the definition. For the other parts there is nothing to prove if $\alpha = \beta$ and otherwise they are easily derived using the definition of $<$. □

The ordering of $\mathbb{R}$ is total:

17.10 Proposition. *Let α and β be real numbers. Then $\alpha \leq \beta$ or $\beta \leq \alpha$.*

Proof. This is a consequence of the remarks made before definition 17.7. □

17.11 Corollary. *Let (a_n) be a Cauchy sequence in $\mathbb{Q}$ with $a_n \geq 0$ for all $n \in \mathbb{N}$. Then for $\alpha = [(a_n)]$ we have $\alpha \geq 0$.*

Proof. Suppose $\alpha < 0$. Then there is a $C \in \mathbb{Q}$ with $C > 0$ and an $n \in \mathbb{N}$ such that $a_n < -C$ for all $n \geq N$. Contradiction. So not $\alpha < 0$. From proposition 17.10 follows that $\alpha \geq 0$. □

17.12 Corollary. *Let $\alpha \in \mathbb{R}$ with $\alpha > 0$. Then there is an $a \in \mathbb{Q}$ with $0 < a < \alpha$.*

Proof. Let (a_n) be a representative of α. Since $\alpha > 0$, there is a $C \in \mathbb{Q}$ and an $N \in \mathbb{N}$ with $C > 0$ and $a_n > C$ for all $n \geq N$. For such an n we have $a_n - C > 0$ and so by corollary 17.11 $\alpha - C \geq 0$. Take for instance $a = \frac{C}{2}$. □

17.1.4 The absolute value on $\mathbb{R}$

The absolute value on $\mathbb{Q}$ can be extended to an absolute value on $\mathbb{R}$. This is needed for the notion of limit in $\mathbb{R}$.

17.13 Definition. Let α be a real number. Then we define the *absolute value* $|\alpha|$ of α as follows

$$|\alpha| = \begin{cases} \alpha & \text{if } \alpha \geq 0 \\ -\alpha & \text{if } \alpha \leq 0. \end{cases}$$

The map $\mathbb{R} \to \mathbb{R}^{\geq 0}$ is called the *absolute value on* $\mathbb{R}$.

17.14 Proposition. *The absolute value $\mathbb{R} \to \mathbb{R}^{\geq 0}$, $\alpha \mapsto |\alpha|$ satisfies the requirements for absolute values: for all $\alpha, \beta \in \mathbb{R}$ we have*

(i) $|\alpha| = 0 \iff \alpha = 0$,
(ii) $|\alpha\beta| = |\alpha||\beta|$,
(iii) $|\alpha + \beta| \leq |\alpha| + |\beta|$.

Proof. See the proof of proposition 7.34. □

The absolute values of real numbers are numbers in $\mathbb{R}^{\geq 0}$. In the previous chapter absolute values of numbers were elements of $\mathbb{Q}^{\geq 0}$. The main reason was that $\mathbb{R}$ still had to be constructed. The absolute value of a number is seen as its distance to 0. Usually distances take values in $\mathbb{R}^{\geq 0}$. Again the function $\mathbb{R} \times \mathbb{R} \to \mathbb{R}^{\geq 0}$, $(\alpha, \beta) \mapsto |\alpha - \beta|$ is a metric, a metric on $\mathbb{R}$. In fact it works for every absolute value that way.

17.15 Proposition. *Let $\alpha \in \mathbb{R}$ be represented by the Cauchy sequence (a_n) in $\mathbb{Q}$. Then*

$$|\alpha| = [(|a_n|)].$$

(So: $|\alpha|$ is the real number represented by the Cauchy sequence $(|a_n|)$.

Proof. If $\alpha = 0$, then (a_n) is a null sequence and so is $(|a_n|)$.

If $\alpha > 0$, then there is an $N \in \mathbb{N}$ such that $a_n > 0$ for $n \geq N$. So $|\alpha| = \alpha = [(a_n)] = [(|a_n|)]$.

If $\alpha < 0$, then there is an $N \in \mathbb{N}$ such that $a_n < 0$ for $n \geq N$. So in this case $|\alpha| = -\alpha = -[(a_n)] = [(-a_n)] = [(|a_n|)]$. □

We could have taken this as a definition of $|\alpha|$ and from that derive the here given definition as a proposition.

17.2 The Completeness of $\mathbb{R}$

On the field $\mathbb{R}$ we have an absolute value which extends the absolute value on $\mathbb{Q}$. So as for $\mathbb{Q}$ we have for $\mathbb{R}$ the notions null sequence, convergent sequence and Cauchy sequence as well. Many Cauchy sequences in $\mathbb{Q}$ do not converge in $\mathbb{Q}$. We will show that Cauchy sequences in $\mathbb{Q}$ do converge in $\mathbb{R}$, which in fact was our objective when constructing $\mathbb{R}$. We will see that even every Cauchy sequence in $\mathbb{R}$ converges. Because of this the field $\mathbb{R}$ is called *complete* with respect to the absolute value on $\mathbb{R}$.

17.16 Proposition. *Let (a_n) be a Cauchy sequence in $\mathbb{Q}$. Then in $\mathbb{R}$ it converges to the real number $[(a_n)]$.*

Proof. Let $\alpha = [(a_n)]$ and let $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$. By Corollary 17.12 there is an $e \in \mathbb{Q}$ with $0 < e < \varepsilon$. The sequence (a_n) is a Cauchy sequence in $\mathbb{Q}$. So there is an $N \in \mathbb{N}$ such that $|a_m - a_n| < e$ for all $m, n \geq N$. For a fixed $m \in \mathbb{N}$ the sequence $(a_m - a_n)$ is a Cauchy sequence in $\mathbb{Q}$ and $[(a_m - a_n)] = a_m - \alpha$. From proposition 17.15 follows that $|a_m - \alpha| = [(|a_m - a_n|)]$. For $m \geq N$ we have $|a_m - a_n| < e$ for all $n \geq N$ and by Corollary 17.11 (applied to $(e - |a_m - a_n|)$) we then have $|a_m - \alpha| \leq e$. So: $|a_m - \alpha| \leq e < \varepsilon$ for all $m \geq N$, that is $\lim_n a_n = \alpha$. □

So: if $\alpha \in \mathbb{R}$ is represented by the Cauchy sequence (a_n) in $\mathbb{Q}$, that is $\alpha = [(a_n)]$, then $\lim_n a_n = \alpha$.

In particular near every real number there is a rational number within any prescribed distance:

17.17 Corollary. *Let α be a real number and let $\varepsilon > 0$. Then there is a rational number a with $|a - \alpha| < \varepsilon$.*

Proof. There is a sequence (a_n) of rational numbers with $\lim_n a_n = \alpha$. So there is an $N \in \mathbb{N}$ with $|a_n - \alpha| < \varepsilon$ for all $n \geq N$. In particular $|a_N - \alpha| < \varepsilon$. □

Next we prove the completeness of $\mathbb{R}$.

17.18 Theorem. *Let (α_n) be a Cauchy sequence of real numbers. Then the sequence (α_n) converges to a real number.*

Proof. For every $n \in \mathbb{N}$ choose a rational number a_n with $|\alpha_n - a_n| < \frac{1}{n+1}$. Then the sequence $(a_n - \alpha_n)$ is a null sequence and so the sequence (a_n) is a Cauchy sequence too: $a_n = \alpha_n + (a_n - \alpha_n)$. The sequence (a_n) converges and so does (α_n). □

In this proof we used that the sum of a null sequence and a Cauchy sequence is a Cauchy sequence. For sequences in $\mathbb{Q}$ we showed this in the previous chapter. That chapter is organized in such a way that for several sections all definitions, lemmas, propositions, theorems and their proofs can be generalized from the rationals to the reals without further ado. Some comments on these sections:

16.2 The notion of null sequence in $\mathbb{R}$ we already used. Also the notion of bounded sequence is applicable to sequences in $\mathbb{R}$. All lemmas and propositions do hold in $\mathbb{R}$. In particular Bernoulli's Inequality holds for $x \in \mathbb{R}$ with $x \geq -1$.

16.3 The notion of convergence we already used. We do have notions of descendence and ascendence for sequences in $\mathbb{R}$. All lemmas and propositions do hold for sequences in $\mathbb{R}$.

16.4 We extend the transformation γ of $\mathbb{Q}$ to $\mathbb{R}$: $\gamma(\alpha) = g\alpha - \lfloor g\alpha \rfloor$ for $\alpha \in \mathbb{R}$. This leads to the base g expansion $(\lfloor \gamma^{n-1}(\alpha) \rfloor)_{n \geq 1}$ of a real number α with $0 \leq \alpha < 1$. The propositions 16.37, 16.38 and 16.40 are about the repetition of the base g expansion of rational numbers.

Simon Stevin (Bruges 1548 – The Hague 1620)

The decimal representation of real numbers and its use in mathematics is introduced by Simon Stevin. He had an important role in organizing the struggle of the Northern Netherlands against the Spanish rule. He wrote books on many subjects: mechanics, hydrostatics, astronomy, the division of the octave in twelve intervals, real numbers, triangular geometry, perspective, algebra, politics. He had a profound understanding of the nature of the real numbers, but the imaginary numbers he could not accept. The Dutch word 'wiskunde' for mathematics originated from Stevin.

16.5 The notion of Cauchy sequence we already used. Completeness of $\mathbb{R}$ means that the notions of Cauchy sequence and convergent sequence coincide. In theorem 16.51 the conclusion is about a sequence being a Cauchy sequence. For the real numbers we thus obtain Cantor's Theorem as formulated below.

Real numbers have g-adic expansions. Let $g \in \mathbb{N}$ with $g \geq 2$. If $(c_n)_{n\geq 1}$is a sequence in $\mathbb{N}_g$, then by Corollary 16.52 the sequence $(a_n)_{n\geq 1}$ with $a_n = \sum_{k=1}^{n} \frac{c_k}{g^k}$ is a Cauchy sequence of rational numbers. This sequence converges in $\mathbb{R}$. Thus we have a correspondence between

$$\text{real numbers } \alpha \text{ with } 0 \leq \alpha < 1$$

and

$$\text{sequences in } \mathbb{N}_g \text{ without } g-1\text{-tail.}$$

Under this correspondence rational numbers correspond to repeating sequences in $\mathbb{N}_g$.

17.19 Theorem (Cantor). *Let (α_n) be an ascending sequence of real numbers and (β_n) a descending sequence of real numbers. Let moreover $\alpha_n \leq \beta_n$ for all $n \in \mathbb{N}$ and that the sequence $(\alpha_n - \beta_n)$ is a null sequence. Then (α_n) and (β_n) converge and for all $m \in \mathbb{N}$*

$$\alpha_m \leq \lim_n \alpha_n = \lim_n \beta_n \leq \beta_m.$$

Proof. As in the proof of theorem 16.51 it follows that (α_n) and (β_n) are Cauchy sequences. By completeness of $\mathbb{R}$ they converge and since they differ by a null sequence their limits are equal. Let $m \in \mathbb{N}$. Since (α_n) is ascending we have $\alpha_n - \alpha_m \geq 0$ for all $n \geq m$. So by Corollary 17.11 $\alpha - \alpha_m \geq 0$: the limit of the sequence (α_n) is greater than or equal to each of its terms. □

Bernard Placidus Johann Nepomuk Bolzano (Prague 1781 – Prague 1848)

Bolzano developed new foundations for analysis. This remained largely unknown till after his death, mainly because he did not submit several of his manuscripts for publication. He gave for example a definition of a Cauchy sequence four years before it appeared in Cauchy's work. Bolzano was the first to use the word set. His ideas on infinity in mathematics anticipated Cantor's theory of infinite sets.

17.20 Example. The sequence (a_n) with

$$a_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots + \frac{1}{n^2}$$

satisfies $a_1 \leq a_2 \leq \ldots$. For the sequence (b_n) with $b_n = a_n + \frac{1}{n}$ we have $b_1 \geq b_2 \geq \ldots$:

$$\begin{aligned} b_n - b_{n+1} &= a_n + \frac{1}{n} - a_{n+1} - \frac{1}{n+1} = \frac{1}{n} - \frac{1}{n+1} - \frac{1}{(n+1)^2} \\ &= \frac{(n+1)^2 - n(n+1) - n}{n(n+1)^2} = \frac{1}{n(n+1)^2}. \end{aligned}$$

Since $b_n - a_n = \frac{1}{n}$ Cantor's Theorem can be applied. Because $a_4 = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} = \frac{205}{144}$ and $b_4 = a_4 + \frac{1}{4} = \frac{205}{144} + \frac{1}{4} = \frac{241}{144}$ the limit λ is located in a segment of length $\frac{1}{4}$, namely $\frac{205}{144} \leq \lambda \leq \frac{241}{144}$. In fact, as Euler showed in 1735, $\lambda = \frac{\pi^2}{6}$.

A direct consequence of Cantor's Theorem is the following.

17.21 Theorem (Bolzano-Weierstraß). *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof. Let (γ_n) be a bounded sequence in $\mathbb{R}$. There is a $C \in \mathbb{R}$ with $|\gamma_n| \leq C$ for all $n \in \mathbb{N}$. We define sequences (α_n) and (β_n) by

$$(\alpha_0, \beta_0) = (-C, C)$$

and for all $n \in \mathbb{N}$:

$$(\alpha_{n+1}, \beta_{n+1}) = \begin{cases} \left(\alpha_n, \dfrac{\alpha_n + \beta_n}{2}\right) & \text{if there are infinitely many } k \in \mathbb{N} \text{ with} \\ & \alpha_n \leq \gamma_k \leq \frac{\alpha_n + \beta_n}{2}, \\ \left(\dfrac{\alpha_n + \beta_n}{2}, \beta_n\right) & \text{otherwise.} \end{cases}$$

Note that for every $n \in \mathbb{N}$ there are infinitely many $k \in \mathbb{N}$ with $\frac{\alpha_n+\beta_n}{2} \leq \gamma_k \leq \beta_n$ if there are only finitely many with $\alpha_n \leq \gamma_k \leq \frac{\alpha_n+\beta_n}{2}$. Now choose a subsequence $(\gamma_{i(n)})$ with $\alpha_n \leq \gamma_{i(n)} \leq \beta_n$ for every $n \in \mathbb{N}$. This subsequence converges by Cantor's Theorem. □

And a consequence of this is:

17.22 Lemma (Weierstraß). *Every bounded ascending sequence in $\mathbb{R}$ converges.*

Proof. Let (α_n) be a bounded ascending sequence in $\mathbb{R}$. Then by theorem 17.21 (α_n) has a convergent subsequence $(\alpha_{i(n)})$ and as in the proof of proposition 16.30 it follows that (α_n) converges. □

The Supremum Property

The field $\mathbb{R}$ is complete: every Cauchy sequence in $\mathbb{R}$ converges. Completeness can be characterized in other ways as well. One of these ways is by the Supremum Property. It is often used, though not so in this book.

17.23 Definition. A $\lambda \in \mathbb{R}$ is called an *upper bound* of a set $X \subseteq \mathbb{R}$ if $x \leq \lambda$ for all $x \in X$. If X has a upper bound, then it is also said that X is *bounded above*. If the set of upper bounds of X has a least element, then this element is called the *least upper bound* or the *supremum* of X. The supremum of X is denoted by $\sup(X)$. The *greatest lower bound* or the *infimum* of X is denoted by $\inf(X)$ (if it exists).

17.24 Examples. The sets

$$\{\, x \in \mathbb{R} \mid 0 < x < 1 \,\},\ \{\, x \in \mathbb{R} \mid 0 \leq x \leq 1 \,\},\ \{\, x \in \mathbb{Q} \mid 0 < x < 1 \,\},$$
$$\text{and } \{\, 1 - \tfrac{1}{n} \mid n \in \mathbb{N}^+ \,\}$$

all four have the supremum 1.

17.25 Theorem (The Supremum Property). *Let X be a nonempty subset of $\mathbb{R}$ which is bounded above. Then X has a supremum.*

Proof. Take an $\alpha \in X$ and an upper bound λ of X. Then $\alpha \leq \lambda$. We define a sequence (α_n, λ_n) in $\mathbb{R}^2$ by $(\alpha_0, \lambda_0) = (\alpha, \lambda)$ and for all $n \in \mathbb{N}$:

$$(\alpha_{n+1}, \lambda_{n+1}) = \begin{cases} \left(\alpha_n, \dfrac{\alpha_n + \lambda_n}{2}\right) & \text{if } \dfrac{\alpha_n + \lambda_n}{2} \text{ is an upper bound,} \\ \left(\dfrac{\alpha_n + \lambda_n}{2}, \lambda_n\right) & \text{otherwise.} \end{cases}$$

Then (α_n) is ascending, (λ_n) descending, $\alpha_n \leq \lambda_n$ for all n and $(\lambda_n - \alpha_n)$ is a null sequence, because $\lambda_n - \alpha_n = \frac{1}{2^n}(\lambda - \alpha)$. By theorem 17.19 the sequences (α_n) and (β_n) converge to a $\mu \in \mathbb{R}$. We will show that μ is the least upper bound of X.

Let $x \in X$. Then $x \leq \lambda_n$ for all n and so $x \leq \lim_n \lambda_n = \mu$.

So μ is an upper bound of X.

Suppose $\mu' < \mu$ and μ' is an upper bound. Since $(\lambda_n - \alpha_n)$ is a null sequence, there is an $n \in \mathbb{N}$ such that $\lambda_n - \alpha_n < \mu - \mu'$. There is an $x \in X$ with $x \geq \alpha_n$. We then have

$$\alpha_n \leq x \leq \mu' < \mu \leq \lambda_n.$$

It follows that $\mu - \mu' \leq \lambda_n - \alpha_n$. Contradiction.

So for every upper bound μ' of X we have $\mu \leq \mu'$. Therefore, μ is the least upper bound. □

We also have that a nonempty set X which bounded below has an infimum: apply the theorem for the set $-X$.

In subsection 17.1.1 we constructed the set $\mathbb{R}$ of real numbers. In subsection 17.1.2 addition and multiplication were defined, giving $\mathbb{R}$ the structure of a field. With the definition of an ordering in subsection 17.1.3 we have given $\mathbb{R}$ the structure of an ordered field. In this section it has been shown that the ordering has the Supremum Property. Note that we started in section 4.2 with Peano's axioms for $\mathbb{N}$ and subsequently constructed $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$. It can be shown that $\mathbb{R}$ is (up to isomorphism) the unique ordered field for which the Supremum Property holds. This makes an axiomatic treatment of $\mathbb{R}$ possible: start with the axioms for an ordered field and add the Supremum Property as an axiom. After that define which numbers are natural, are integral or are rational. This approach is often taken in textbooks for mathematical analysis: it gives a short route to calculus.

17.3 Convergence of Series

17.26 Theorem. *Let $\sum_{n=0}^{\infty} a_n$ be a series of real numbers such that the series $\sum_{n=0}^{\infty} |a_n|$ converges. Then the series $\sum_{n=0}^{\infty} a_n$ converges.*

Proof. The series $\sum_{n=0}^{\infty} |a_n|$ is a Cauchy sequence and from

$$\left| \sum_{k=m}^{n-1} a_k \right| \leq \sum_{k=m}^{n-1} |a_k|$$

follows that $\sum_{n=0}^{\infty} a_n$ is also a Cauchy sequence. □

17.27 Definition. A series $\sum_{n=0}^{\infty} a_n$ is called an *absolute* convergent series if the series $\sum_{n=0}^{\infty} |a_n|$ converges.

More generally:

17.28 Theorem. *Let $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ be series of real numbers and let $|a_n| \leq b_n$ for all $n \in \mathbb{N}$. Suppose that the series $\sum_{n=0}^{\infty} b_n$ converges. Then the series $\sum_{n=0}^{\infty} a_n$ is absolutely convergent.*

Proof. This follows from

$$\left|\sum_{k=m}^{n-1} a_k\right| \leq \sum_{k=m}^{n-1} |a_k| \leq \sum_{k=m}^{n-1} b_k \qquad \square$$

17.29 Terminology. For $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ as in theorem 17.28 one says that the series $\sum_{n=0}^{\infty} a_n$ is *majorized* by the convergent series $\sum_{n=0}^{\infty} b_n$.

17.30 Example. For $t \in \mathbb{N}$ with $m \geq 2$ the series $\sum_{n=1}^{\infty} \frac{1}{n^t}$ converges since it is majorized by the convergent series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ (example 17.20).

17.4 Polynomial Equations over $\mathbb{R}$

Let $f(x)$ be a monic polynomial of degree m:

$$f(x) = x^m + \alpha_1 x^{m-1} + \cdots + \alpha_k x^{m-k} + \cdots + \alpha_m, \tag{17.1}$$

where $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$. (*Monic* means that the leading coefficient is 1.) We study the question whether the equation $f(x) = 0$ has a solution, that is whether the polynomial function $x \mapsto f(x)$ has a zero.

We will use the continuity of polynomial functions. The notion of continuity we use here was introduced by Heine. The usual definition however is Cauchy's.

17.31 Definition. Let $U \subseteq \mathbb{R}$. A function $g\colon U \to \mathbb{R}$ is called *continuous* in $\gamma \in U$ if

$$\lim_n g(\gamma_n) = g(\gamma)$$

for all sequences (γ_n) in U converging to γ. The function $g\colon U \to \mathbb{R}$ is called *continuous* if it is continuous in all $\gamma \in U$.

> We compare the notions of continuity of Heine and Cauchy. In this book only Heine's definition is used.
>
> **Definition (Cauchy).** Let $U \subseteq \mathbb{R}$. A function $g\colon U \to \mathbb{R}$ is called *continuous* in $\gamma \in U$ if for every $\varepsilon > 0$ a $\delta > 0$ exists such that $|g(x) - g(\gamma)| < \varepsilon$ for all $x \in U$ with $|x - \gamma| < \delta$.
>
> **Proposition.** Let $U \subseteq \mathbb{R}$. A function $g\colon U \to \mathbb{R}$ is continuous (in the sense of Heine) in $\gamma \in U$ if and only if it is so in the sense of Cauchy.

H. Eduard Heine (Berlin 1821 – Halle 1881)

Heine was a pupil of Dirichlet. He contributed to the understanding of continuity in mathematics and is especially known by the Heine Borel Theorem, which is about bounded subsets of $\mathbb{R}$ (or, more generally, $\mathbb{R}^n$), closed under taking limits.

PROOF. Suppose g is continuous in γ in the sense of Cauchy. Let (γ_n) be a sequence in U converging to γ. To prove that $(g(\gamma_n))$ converges to $g(\gamma)$. Let $\varepsilon > 0$. There is a $\delta > 0$ with $|g(x) - g(\gamma)| < \varepsilon$ for all $x \in U$ with $|x - u| < \delta$. Since (γ_n) converges to γ, there is an $N \in \mathbb{N}$ such that $|\gamma_n - \gamma| < \delta$ for all $n \in \mathbb{N}$ with $n \geq N$. For this n we have $|g(\gamma_n) - g(\gamma)| < \varepsilon$. So $(g(\gamma_n))$ converges to $g(\gamma)$.

Suppose g is not continuous in γ in the sense of Cauchy. To prove that there is in U a sequence (γ_n) converging to γ while $(g(\gamma_n))$ does not converge to $g(\gamma)$. There is an $\varepsilon > 0$ such that for all $\delta > 0$ there is an $x \in U$ with $|x-\gamma| < \delta$ and $|g(x)-g(\gamma)| \geq \varepsilon$. Take for every $n \in \mathbb{N}^+$ a $\gamma_n \in U$ with $|\gamma_n - \gamma| < \frac{1}{n}$ and $|g(\gamma_n) - g(\gamma)| \geq \varepsilon$. □

17.32 Proposition. *Polynomial functions on $\mathbb{R}$ are continuous.*

PROOF. This is a consequence of the rules for limits: let $f(x)$ be as in (17.1) and suppose that $\lim_n \gamma_n = \gamma$, then

$$\lim_n f(\gamma_n) = \lim_n \sum_{k=0}^{m} \alpha_k \gamma_n^{m-k} = \sum_{k=0}^{m} \lim_n (\alpha_k \gamma_n^{m-k}) = \sum_{k=0}^{m} \alpha_k \gamma^{m-k} = f(\gamma). \qquad \square$$

The next theorem gives conditions for the existence of a zero of a real function $g(x)$.

17.33 Theorem. *Let α and β be real numbers with $\alpha < \beta$. Let $g\colon \mathbb{R} \to \mathbb{R}$ be continuous in all γ with $\alpha < \gamma < \beta$. Suppose that $g(\alpha) < 0$ and $g(\beta) > 0$. Then there is a γ with $\alpha < \gamma < \beta$ and $g(\gamma) = 0$.*

PROOF. We define sequences (α_n) and (β_n) by $(\alpha_0, \beta_0) = (\alpha, \beta)$ and for all $n \in \mathbb{N}$:

$$(\alpha_{n+1}, \beta_{n+1}) = \begin{cases} \left(\alpha_n, \dfrac{\alpha_n + \beta_n}{2}\right) & \text{if } g\left(\dfrac{\alpha_n + \beta_n}{2}\right) \geq 0, \\ \left(\dfrac{\alpha_n + \beta_n}{2}, \beta_n\right) & \text{if } g\left(\dfrac{\alpha_n + \beta_n}{2}\right) < 0. \end{cases}$$

Then by Cantor's Theorem (α_n) and (β_n) converge to a real number γ with $\alpha_n \leq \gamma \leq \beta_n$ for all $n \in \mathbb{N}$. Since g is continuous in γ we have $g(\gamma) = \lim_n g(\alpha_n)$ and so $g(\gamma) \leq 0$, because $g(\alpha_n) < 0$ for all $n \in \mathbb{N}$. Similarly from $g(\beta_n) \geq 0$ for all $n \in \mathbb{N}$ follows $g(\gamma) \geq 0$. So $g(\gamma) = 0$. □

Two consequences of this theorem:

17.34 Corollary. *Let $f(x)$ be a polynomial over $\mathbb{R}$ of odd degree. Then $f(x)$ has a zero in $\mathbb{R}$.*

Proof. We assume the polynomial to be monic. Let $f(x) = x^m + \alpha_1 x^{m-1} + \cdots + \alpha_k x^{m-k} + \cdots + \alpha_m$ where $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ and m odd. From $\lim_n \frac{f(n)}{n^m} = 1$ follows that there exists an $N \in \mathbb{N}$ with $\frac{f(N)}{N^m} > 0$ and so also $f(N) > 0$. From $\lim_n \frac{f(-n)}{(-n)^m} = 1$ follows that there is an $M \in \mathbb{N}$ with $\frac{f(-M)}{(-M)^m} > 0$ and so $f(-M) < 0$, because $(-M)^m < 0$. By proposition 17.32 and theorem 17.33 there is a zero γ of $f(x)$ with $-M < \gamma < N$. □

In $\mathbb{Q}$ the possibility of extracting roots was very limited. In $\mathbb{R}$ this is different.

17.35 Corollary. *Let α be a positive real number and let $m \in \mathbb{N}^+$. Then there is a positive real number γ with $\gamma^m = \alpha$.*

Proof. Consider the polynomial $f(x) = x^m - \alpha$. There is an $N \in \mathbb{N}$ such that $f(N) > 0$ and we have $f(0) = -\alpha < 0$. So there is a γ with $0 < \gamma < N$ and $f(\gamma) = 0$, that is $\gamma^m = \alpha$. □

17.36 Definition. Let $\alpha \geq 0$ and let $m \in \mathbb{N}^+$. The unique real number $\gamma \geq 0$ with $\gamma^m = \alpha$ is called the *m-th root of α*. Notation: $\gamma = \sqrt[m]{\alpha}$.

> The γ is unique since the roots of the equation $x^n - \alpha = 0$ are of the form $\zeta\gamma$ with ζ a root of unity in $\mathbb{R}$, that is $\zeta = 1$ or $\zeta = -1$, see the last paragraph of chapter 13 on page 265.

We have seen that for many polynomial equations there are solutions in $\mathbb{R}$. In particular many equations over $\mathbb{Q}$ having solutions in $\mathbb{R}$ do not have solutions in $\mathbb{Q}$.

17.37 Definition. A number which is a solution of a nontrivial polynomial equation with rational coefficients is called *algebraic*. If a number is not algebraic, then it is called *transcendental*.

We will see that in $\mathbb{R}$ there are many many transcendental numbers.

17.5 Real Numbers and Geometry

As is clear from the previous section square roots of positive reals do exist. In geometry these square roots are needed for the Theorem of Pythagoras.

17.5.1 The number π

For approximating π as done by Archimedes square roots are used.

> The number π is an example of a transcendental number. Here we will not prove this. In 1882 it was the German mathematician **Lindemann** who was the first to give a proof. A consequence of the transcendence of π is the impossibility of 'squaring' the circle: it is impossible starting from (the radius of) a circle to construct by ruler and compass a square of the same area. Thus the 'squaring of the circle', a problem that goes back to Greek antiquity, was solved in a negative sense.

The number π is the ratio of the circumference to the diameter of the circle. The length of a curve is in some way defined as a limit. The idea is approximating a curve by straight line segments and if the curve is not too wild, its length is a limit of sums of lengths of line segments. That is what Archimedes did. He approximated the circle by inscribed and circumscribed regular polygons. In fact he took regular $3 \cdot 2^{n-1}$-gons: a triangle, a hexagon, a 12-gon and so on, inscribed and circumscribed. For $n \in \mathbb{N}^+$ we take a_n to be equal to half the circumference of an inscribed regular $3 \cdot 2^{n-1}$-gon and b_n half the circumference of circumscribed regular $3 \cdot 2^{n-1}$-gon. Using elementary plane geometry it is easily deduced that for all $n \in \mathbb{N}^+$

$$\begin{cases} a_{n+1} = a_n \sqrt{\dfrac{2b_n}{a_n + b_n}}, \\ b_{n+1} = \dfrac{2a_n b_n}{a_n + b_n}. \end{cases}$$

These formulas, together with $a_1 = \frac{3}{2}\sqrt{3}$ and $b_1 = 3\sqrt{3}$, define the pairs a_n, b_n inductively. Theorem 17.19 applies. Geometrically this is clear, however, we derive this from the formulas.

a) $a_n < b_n$ for all n. This follows from

$$\frac{a_{n+1}^2}{b_{n+1}^2} = \frac{\frac{2a_n^2 b_n}{a_n + b_n}}{\frac{4a_n^2 b_n^2}{(a_n + b_n)^2}} = \frac{a_n + b_n}{2b_n} = \frac{\frac{a_n}{b_n} + 1}{2}.$$

b) (b_n) is descending:

$$\frac{b_{n+1}}{b_n} = \frac{2a_n}{a_n + b_n} = \frac{2}{1 + \frac{b_n}{a_n}} < 1.$$

c) (a_n) is ascending:

$$\frac{a_{n+1}^2}{a_n^2} = \frac{2b_n}{a_n + b_n} = \frac{2}{\frac{a_n}{b_n} + 1} > 1.$$

Archimedes (Syracuse 287 BC – Syracuse 212 BC)

Using new methods Archimedes of Syracuse had made advances in Greek geometry. His methods of approximating numbers were already in the spirit of Newton and Leibniz in the seventeenth century. His mathematical precision was not surpassed until in the nineteenth century the notion of limit obtained a solid basis. Archimedes also was very inventive in designing all kinds of instruments, partly for warfare in order to keep the Romans out of Sicily.

d) $(b_n - a_n)$ is a null sequence:

$$b_{n+1}^2 - a_{n+1}^2 = \frac{4a_n^2 b_n^2}{(a_n + b_n)^2} - \frac{2a_n^2 b_n}{a_n + b_n} = \frac{2a_n^2 b_n (b_n - a_n)}{(a_n + b_n)^2}$$

and so

$$b_{n+1} - a_{n+1} = \frac{2a_n^2 b_n (b_n - a_n)}{(a_n + b_n)^2 (a_{n+1} + b_{n+1})} = \frac{2\frac{b_n}{a_n}(b_n - a_n)}{(1 + \frac{b_n}{a_n})^2 (\frac{a_{n+1}}{a_n} + \frac{b_{n+1}}{a_n})}$$
$$< \frac{2\frac{b_1}{a_1}}{8}(b_n - a_n) = \frac{1}{2}(b_n - a_n).$$

Archimedes computed, using $a_6 < \pi < b_6$, that $\frac{223}{71} < \pi < \frac{22}{7}$. For this he had to find approximations of the square roots that occur in this computation. Thus the estimation of Archimedes is obtained using the circumferences of an inscribed and a circumscribed regular 96-gon.

> The use of the letter π for the ratio of the circumference to the diameter of the circle dates from the seventeenth century. In 1647 **Oughtred** denoted the ratio of diameter to circumference as d/π and **Gregory** used π/r for the ratio of circumference to radius. In 1706 **William Jones** denoted the ratio of circumference to diameter, as we do now, as π. Euler adopted this usage in 1737, which contributed a lot to the general acceptance of the use of π.

17.5.2 Coordinates

Descartes introduced the use of coordinates in geometry and thus the start was made of the algebraization of geometry. The plane is identified with $\mathbb{R}^2$, the set

of ordered pairs of real numbers. Geometric notions are translated into algebraic ones. For example the notion of distance. Using a perpendicular coordinate axis it follows that the following definition of distance coincides with the geometric notion.

17.38 Definition. The *distance* $d((x_1, y_1), (x_2, y_2))$ between (x_1, y_1) and (x_2, y_2) is defined by

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

It is geometrically clear that the distance in $\mathbb{R}^2$ is a metric. Here we deduce it from the given definition.

17.39 Proposition. *The distance* $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^{\geq 0}$ *is a metric.*

Proof. Clearly $d((x_1, y_1), (x_2, y_2)) = 0 \iff (x_1, y_1) = (x_2, y_2)$ and also $d((x_1, y_1), (x_2, y_2)) = d((x_2, y_2), (x_1, y_1))$ for all $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$.

We prove the triangle inequality

$$d((x_1, y_1), (x_3, y_3)) \leq d((x_1, y_1), (x_2, y_2)) + d((x_2, y_2), (x_3, y_3)).$$

Write $a_1 = x_1 - x_2$, $a_2 = x_2 - x_3$, $b_1 = y_1 - y_2$ and $b_2 = y_2 - y_3$. Then to prove

$$\sqrt{(a_1 + a_2)^2 + (b_1 + b_2)^2} \leq \sqrt{a_1^2 + b_1^2} + \sqrt{a_2^2 + b_2^2}.$$

This is equivalent to consecutively

$$\begin{aligned}
(a_1 + a_2)^2 + (b_1 + b_2)^2 &\leq a_1^2 + b_1^2 + a_2^2 + b_2^2 + 2\sqrt{(a_1^2 + b_1^2)(a_2^2 + b_2^2)}, \\
a_1 a_2 + b_1 b_2 &\leq \sqrt{(a_1^2 + b_1^2)(a_2^2 + b_2^2)}, \\
(a_1 a_2 + b_1 b_2)^2 &\leq (a_1^2 + b_1^2)(a_2^2 + b_2^2), \\
2a_1 a_2 b_1 b_2 &\leq a_1^2 b_2^2 + a_2^2 b_1^2, \\
0 &\leq (a_1 b_2 - a_2 b_1)^2.
\end{aligned}$$

□

17.6 The Group $\mathbb{R}^*$

For $\mathbb{Q}$ there is a connection between multiplication and addition: by the Fundamental Theorem of Arithmetic multiplication in $\mathbb{Q}^+$ becomes addition of the valuations. We will show that multiplication in $\mathbb{R}^+$ corresponds to addition in $\mathbb{R}$: a group isomorphism from the additive group $\mathbb{R}$ to the multiplicative group $\mathbb{R}^+$. The isomorphism is given by the exponential function. First we introduce the exponential function.

The exponential function

17.40 Definition. Let $x \in \mathbb{R}$. The sequence $(e_n(x))$ in $\mathbb{R}$ is defined by:

$$e_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!} \quad \text{for all } n \in \mathbb{N}.$$

So $(e_n(x))$ is the series having $\frac{x^n}{n!}$ as general term. We will show that this series converges. The limit will be the value of the exponential function in x.

17.41 Lemma. *The sequence $(e_n(x))$ converges for all $x \in \mathbb{R}$.*

Proof. Fix $x \in \mathbb{R}$. It suffices to prove that the series $\sum_{n=N}^{\infty} \frac{x^n}{n!}$ converges for some $N \in \mathbb{N}$. Take $N = \lfloor |x| \rfloor$. For $n \geq N$ put $k = n - N$. We have

$$\Big|\frac{x^n}{n!}\Big| = \frac{|x|^N}{N!} \cdot \frac{|x|^k}{(N+1)(N+2)\dots(N+k)} \leq \frac{|x|^N}{N!}\Big(\frac{|x|}{N+1}\Big)^k.$$

So the series $\sum_{n=N}^{\infty} \frac{x^n}{n!} = \sum_{k=0}^{\infty} \frac{x^{N+k}}{(N+k)!}$ is majorized by a convergent geometric series. □

17.42 Definition. For every $x \in \mathbb{R}$ we define

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

The function $\exp\colon \mathbb{R} \to \mathbb{R}$ is called the *exponential function*.

17.43 Theorem. *For all $x, y \in \mathbb{R}$ we have* $\exp(x+y) = \exp(x)\exp(y)$.

Proof.

$$\begin{aligned} e_n(x)e_n(y) &= \Big(\sum_{k=0}^{n} \frac{x^n}{k!}\Big)\Big(\sum_{k=0}^{n} \frac{y^n}{k!}\Big) = \sum_{0 \leq k,l \leq n} \frac{x^k y^l}{k!l!} \\ &= \sum_{t=0}^{n} \sum_{k=0}^{t} \frac{x^k y^{t-k}}{k!(t-k)!} + \sum_{t=n+1}^{2n} \sum_{k=0}^{n} \frac{x^k y^{t-k}}{k!(t-k)!}. \end{aligned}$$

We have

$$\sum_{t=0}^{n} \sum_{k=0}^{t} \frac{x^k y^{t-k}}{k!(t-k)!} = \sum_{t=0}^{n} \frac{1}{t!} \sum_{k=0}^{t} \binom{t}{k} x^k y^{t-k} = \sum_{t=0}^{n} \frac{(x+y)^t}{t!} = e_n(x+y).$$

Let $z = \max(|x|, |y|)$. Then

$$|e_n(x)e_n(y) - e_n(x+y)| \leq \sum_{t=n+1}^{2n} \sum_{k=0}^{t} \frac{z^t}{k!(t-k)!} = \sum_{t=n+1}^{2n} \frac{z^t}{t!} \sum_{k=0}^{t} \binom{t}{k}$$

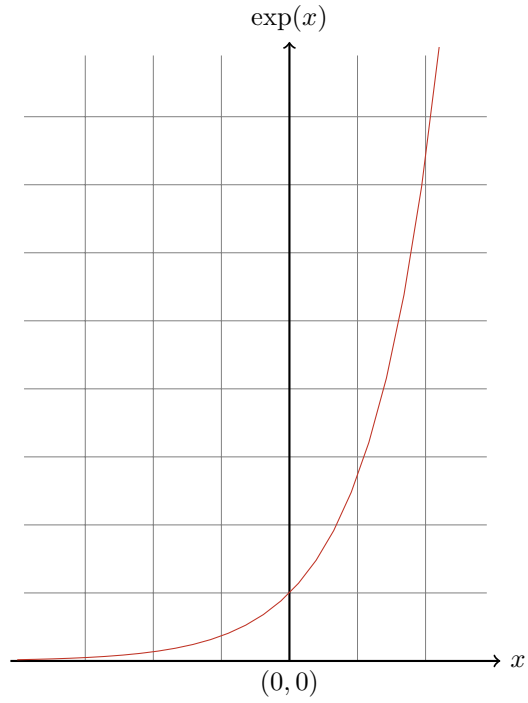


Figure 17.1: graph of the function $\exp\colon \mathbb{R} \to \mathbb{R}^+$

$$= \sum_{t=n+1}^{2n} \frac{z^t}{t!} \cdot 2^t = e_{2n}(2z) - e_n(2z)$$

and $\lim_n(e_{2n}(2z) - e_n(2z)) = 0$ since $(e_n(2z))$ converges. □

17.44 Corollary. *For all $x \in \mathbb{R}$ we have $\exp(x) > 0$.*

Proof. For $x \geq 0$ it is clear from the definition that $\exp(x) \geq 1$. From $\exp(-x)\exp(x) = \exp(0) = 1$ it follows that $0 < \exp(x) \leq 1$ for $x \leq 0$. □

17.45 Corollary. *Let x_1 and x_2 be real numbers with $x_1 < x_2$. Then $\exp(x_1) < \exp(x_2)$.*

Proof. $\exp(x_2) = \exp(x_2 - x_1 + x_1) = \exp(x_2 - x_1)\exp(x_1) > \exp(x_1)$. □

The function exp can be seen as a map $\exp\colon \mathbb{R} \to \mathbb{R}^+$. We will show it is a bijection.

17.46 Theorem. *The exponential function is continuous.*

Proof. First we prove that exp is continuous in 0. Let (γ_n) be a null sequence. To prove that $\exp(\gamma_n)$ converges to $\exp(0) = 1$, that is $(\exp(\gamma_n) - 1)$ is a null sequence. Let $N \in \mathbb{N}$ be such that $|\gamma_n| < \frac{1}{2}$ for all $n \geq N$. Then for $n \geq N$:

$$|\exp(\gamma_n) - 1| = \left| \sum_{k=1}^{\infty} \frac{\gamma_n^k}{k!} \right| \leq \sum_{k=1}^{\infty} |\gamma_n|^k = \frac{|\gamma_n|}{1 - |\gamma_n|} < 2|\gamma_n|.$$

Since (γ_n) is a null sequence, $(\exp(\gamma_n) - 1)$ is a null sequence as well.

Now let (γ_n) be any convergent sequence, say with limit γ. Then

$$\lim_n \exp(\gamma_n) = \lim_n \exp(\gamma_n - \gamma) \exp(\gamma) = \exp(\gamma). \qquad \square$$

17.47 Corollary. *Let $y \in \mathbb{R}$ with $y > 0$. Then there exists an $x \in \mathbb{R}$ with $\exp(x) = y$.*

Proof. Let $y \in \mathbb{R}$ with $y > 1$. Since the function exp is continuous in every real number, so is the function $g\colon x \mapsto \exp(x) - y$. We have $g(0) = 1 - y < 0$. Take $z > y - 1$. Then $g(z) = \exp(z) - y > 1 + z - y > 0$. Because g is continuous in every number, it follows from theorem 17.31 that there exists an x with $0 < x < z$ such that $g(x) = 0$, that is $\exp(x) = y$. For $0 < y < 1$: there is an x such that $\exp(x) = \frac{1}{y}$, that is $\exp(-x) = y$. $\square$

17.48 Theorem. *The map $\exp\colon \mathbb{R} \to \mathbb{R}^+$ is an isomorphism of groups.*

Proof. The map exp is injective by Corollary 17.45. By Corollary 17.47 the map exp is surjective. So by theorem 17.43 we see that exp is an isomorphism from the group $\mathbb{R}$ (under addition) to the group $\mathbb{R}^+$ (under multiplication). $\square$

17.49 Corollary. *The map*

$$\mathbb{Z}/2 \times \mathbb{R} \to \mathbb{R}^*, \quad (\overline{k}, x) \mapsto (-1)^k \exp(x)$$

is an isomorphism of groups.

Proof. It is a composition of group isomorphisms:

$$\mathbb{Z}/2 \times \mathbb{R} \xrightarrow{\sim} \mathbb{Z}/2 \times \mathbb{R}^+ \xrightarrow{\sim} \mathbb{R}^*,$$

where the last map is given by $(\overline{k}, x) \mapsto (-1)^k x$. $\square$

Thus multiplication in $\mathbb{R}^*$ is translated into addition in $\mathbb{Z}/2 \times \mathbb{R}$. Since $\exp\colon \mathbb{R} \to \mathbb{R}^+$ is bijective it has an inverse:

17.50 Definition. The function $\log\colon \mathbb{R}^+ \to \mathbb{R}$ is defined as the inverse of the function exp. The function log is called the *logarithm* or also the *natural logarithm* and is also denoted by: ln.

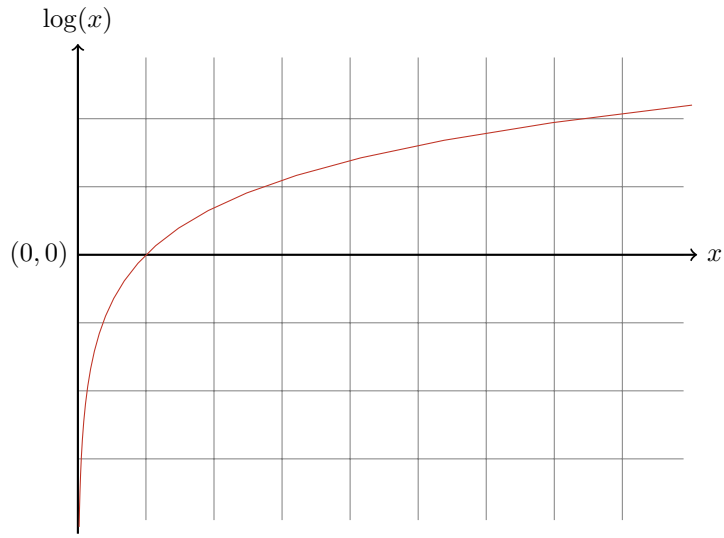


Figure 17.2: graph of the function $\log\colon \mathbb{R}^+ \to \mathbb{R}$

We have seen that for every $\beta > 0$ and every $m \in \mathbb{N}^+$ there is a $\gamma > 0$ with $\gamma^m = \beta$. Another proof is as follows:

$$(\log(\tfrac{1}{m}\exp(\beta)))^m = \log(\exp(\beta)) = \beta.$$

Extraction of the m-th root from β is translated into division of $\exp(\beta)$ by m.

Now we can define β^x for all $\beta \geq 0$ and all $x \in \mathbb{R}$:

17.51 Definition. Let β and x be real numbers with $\beta > 0$. We define β *to the power* x:

$$\beta^x = \exp(x \log \beta).$$

The function $\mathbb{R} \to \mathbb{R}$, $x \mapsto \beta^x$ is called the *exponential function with base* β.

For $x \in \mathbb{Z}$ this is not new: $\exp(x \log \beta) = \exp(\log(\beta^x)) = \beta^x$. For $x = \frac{1}{m}$ with $m \in \mathbb{N}^+$ we have $\beta^{\frac{1}{m}} = \sqrt[m]{\beta}$.

17.52 Theorem. *Let β, x and y be real numbers with $\beta > 0$. Then*

(i) $\beta^{x+y} = \beta^x \beta^y$,
(ii) $(\beta^x)^y = \beta^{xy}$.

Proof.

(i) $\beta^x \beta^y = \exp(x \log \beta) \exp(y \log \beta) = \exp((x + y) \log \beta) = \beta^{x+y}$.
(ii) From $\beta^x = \exp(x \log \beta)$ it follows that $\log \beta^x = x \log \beta$ for all $x \in \mathbb{R}$ and so $(\beta^x)^y = \exp(y \log \beta^x) = \exp(xy \log \beta) = \beta^{xy}$. □

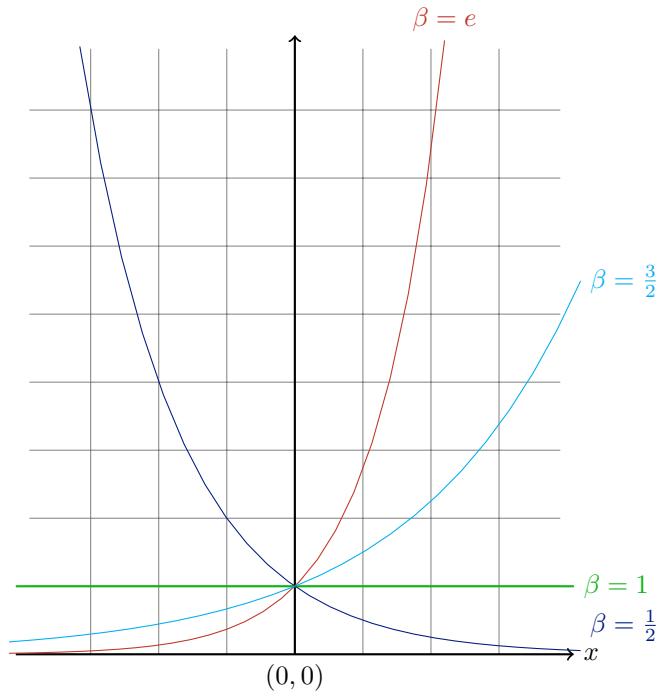


Figure 17.3: graphs of functions $x \mapsto \beta^x$ for $\beta = \frac{1}{2}, 1, \frac{3}{2}, e$

17.53 Notation.

Thus the function exp is the exponential function with base e:

$$e^x = \exp(x \log e) = \exp(x).$$

Interest in issues concerning the number e dates back to the seventeenth century (**Napier**, **Briggs**, **Huygens**, **Mercator**). It was mostly about exponential functions and not explicitly about e. Jacob Bernoulli studied the number e as the limit of the sequence $((1+\frac{1}{n})^n)$ in relation to compound interest without relating it to the exponential function. **Leibniz** was the first to use a special notation for the number e, namely b. He did so in a letter to Huygens in 1690. The use of the letter e started with Euler. Possibly he used the letter e because he wanted to use a vowel and the a was already in use. He computed the approximation 2.718281828459045235. The first 20 terms of $\sum_{n=0}^{\infty} \frac{1}{n!}$ are needed for this. He also proved e to be the limit of the sequence $((1 + \frac{1}{n})^n)$.

Charles Jean Gustave Nicola Baron de la Vallée Poussin (Louvain 1866 – Louvain 1962)
Jacques Salomon Hadamard (Versailles 1865 – Paris 1963)

The Belgian Charles de la Vallée Poussin (left) and the French Jacques Hadamard (right) independently proved the Prime Number Theorem. De la Vallée Poussin is also known for his work *Cours d'Analyse* on mathematical analysis. Hadamard published on differential equations and stochastics.

Gauß conjectured (in 1791, being fourteen years of age) that a good approximation of $\pi(n)$, see definition 15.1, is given by $\frac{n}{\log n}$, or more precisely

$$\lim_n \frac{\pi(n)\log n}{n} = 1.$$

This conjecture was proved more than a century later, in 1896, by **Hadamard** and **de la Vallée Poussin**. Now this is known as the *Prime Number Theorem*.

Powers

The exponential function with base $\beta > 0$ is a map $\mathbb{R} \to \mathbb{R}^+$, $x \mapsto \beta^x$. If in β^x we fix x and let β vary in $\mathbb{R}^+$ we get a transformation of $\mathbb{R}^+$.

17.54 Definition. Let $\alpha \in \mathbb{R}$. The *power function* $\mathbb{R}^+ \to \mathbb{R}^+ : x \mapsto x^\alpha$ raises x to the power α. It generalizes the familiar m-th power function $x \mapsto x^m$ for $m \in \mathbb{Z} \setminus \{0\}$. See Figure 17.4.

For $m \in \mathbb{N}^+$, $k \in \mathbb{Z}$ and $\alpha \in \mathbb{R}^+$ we have

$$((-1)^k\alpha)^m = (-1)^{km}\alpha^m \in \mathbb{R}^+.$$

So for m odd all the elements of $\mathbb{R}^*$ are m-th powers, whereas for m even it are the elements of $\mathbb{R}^+$. In particular for $m = 2$ the group $\mathbb{R}^*$ is the disjoint union of $\mathbb{R}^+$ and $-\mathbb{R}^+$; the first set consists of all squares and the second of all nonsquares.

17.7 Infinite Continued Fractions

We have seen how with the use of Euclid's algorithm a rational number can be written as $\langle a_1, \ldots, a_n \rangle$, the continued fraction of numbers $a_1 \in \mathbb{Z}$ and $a_2, \ldots, a_n \in$

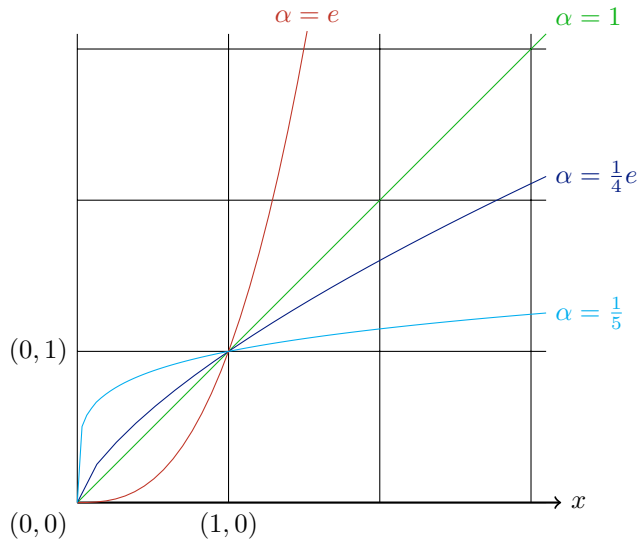


Figure 17.4: the power function $x \mapsto x^\alpha$ for $\alpha = \frac{1}{5}, \frac{1}{4}e, 1, e$

$\mathbb{N}^+$. When applied to a rational number r, this process produces:

$$r = \langle a_1, r_2 \rangle = \langle a_1, a_2, r_3 \rangle = \cdots = \langle a_1, \ldots, a_{n-1}, r_n \rangle,$$

where (with $r_1 = r$):

$$\begin{cases} a_k = \lfloor r_k \rfloor \\ r_{k+1} = \dfrac{1}{r_k - a_k} \end{cases}$$

for $k = 1, \ldots, n-1$. The process ends as soon as $r_n \in \mathbb{Z}$.

This can be applied to an irrational number α as well. Thus we obtain:

$$\begin{aligned} \alpha &= \langle \lfloor \alpha \rfloor, \varphi(\alpha) \rangle = \langle \lfloor \alpha \rfloor, \lfloor \varphi(\alpha) \rfloor, \varphi^2(\alpha) \rangle = \cdots \\ &= \langle \lfloor \alpha \rfloor, \lfloor \varphi(\alpha) \rfloor, \ldots, \lfloor \varphi^{n-1}(\alpha) \rfloor, \varphi^n(\alpha) \rangle, \end{aligned}$$

where φ is the transformation of $\mathbb{R} \setminus \mathbb{Q}$ mapping an irrational number α to $\frac{1}{\alpha - \lfloor \alpha \rfloor}$, that is, $\varphi(\alpha)$ is defined by

$$\alpha = \lfloor \alpha \rfloor + \frac{1}{\varphi(\alpha)}.$$

Infinite continued fractions provide good approximations of irrational numbers by rational ones. How good such an approximation is, is the subject of the next section.

17.55 Example. We take $\alpha = \sqrt{2}$. Then we get:

$$\begin{aligned}
\sqrt{2} &= 1 + (\sqrt{2} - 1) \\
\varphi(\sqrt{2}) &= \frac{1}{\sqrt{2} - 1} = \sqrt{2} + 1 = 2 + (\sqrt{2} - 1) \\
\varphi^2(\sqrt{2}) &= \frac{1}{\sqrt{2} - 1} = \sqrt{2} + 1 = 2 + (\sqrt{2} - 1) \\
&\vdots
\end{aligned}$$

So:

$$\begin{aligned}
\sqrt{2} &= \langle 1, \sqrt{2} + 1 \rangle = \langle 1, 2, \sqrt{2} + 1 \rangle = \langle 1, 2, 2, \sqrt{2} + 1 \rangle = \cdots \\
&= \langle 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, \sqrt{2} + 1 \rangle.
\end{aligned}$$

We are tempted to write

$$\sqrt{2} = \langle 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, \dots \rangle,$$

but for that we first have to assign a meaning to the right hand side and then it still has to be shown that we have here an equality.

17.56 Theorem. *Let $a_1, a_2, a_3, \dots$ be a sequence with $a_1 \in \mathbb{Z}$ and $a_2, a_3, \dots \in \mathbb{N}^+$. Then the sequence $r_1, r_2, r_3, \dots$ of rational numbers defined by*

$$r_n = \langle a_1, \dots, a_n \rangle \qquad \textit{(for all } n \in \mathbb{N}^+\textit{)}$$

converges.

Proof. We compute $r_{n+1} - r_n$:

$$\begin{aligned}
r_{n+1} - r_n &= \langle a_1, \dots, a_{n+1} \rangle - \langle a_1, \dots, a_n \rangle = \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \\
&= \frac{p_{n+1}q_n - p_n q_{n+1}}{q_n q_{n+1}} = \frac{(-1)^{n+1}}{q_n q_{n+1}}.
\end{aligned}$$

From the definition of the numbers $q_1, q_2, \dots$ it easily follows that

$$1 = q_1 \leq q_2 < q_3 < q_4 < \cdots .$$

The sequence of the differences $r_{n+1} - r_n$ is alternating positive and negative and the sequence $(|r_{n+1} - r_n|)$ descends with 0 as limit. Hence the sequence (r_n) converges. □

17.57 Definition. For numbers $a_1, a_2, a_3, \dots$ with $a_1 \in \mathbb{Z}$ and $a_2, a_3, \dots \in \mathbb{N}^+$ we define the (infinite) *continued fraction* of $a_1, a_2, a_3, \dots$ by

$$\langle a_1, a_2, a_3, \dots \rangle = \lim_n \langle a_1, a_2, \dots, a_n \rangle.$$

17.58 Definition. Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. The sequence $\lfloor \alpha \rfloor, \lfloor \varphi(\alpha) \rfloor, \lfloor \varphi^2(\alpha) \rfloor, \dots$ is called the *continued fraction expansion* of α.

17.59 Example. We have seen that $1, 2, 2, 2, 2, 2, 2, 2, \dots$ is the continued fraction expansion of $\sqrt{2}$. The numbers $\langle 1, 2, 2, 2, 2, \dots, 2 \rangle$ can be computed as in the previous section:

i:	-1	0	1	2	3	4	5	6	7	8	$\cdots$
a_i:	–	–	1	2	2	2	2	2	2	2	$\cdots$
p_i:	0	1	1	3	7	17	41	99	239	577	$\cdots$
q_i:	1	0	1	2	5	12	29	70	169	408	$\cdots$

So we have:

$$1 < \tfrac{7}{5} < \tfrac{41}{29} < \cdots < \langle 1, 2, 2, 2, 2, 2, 2, 2, 2, \dots \rangle < \cdots < \tfrac{99}{70} < \tfrac{17}{12} < \tfrac{3}{2}.$$

We will show that the continued fraction given by the continued fraction expansion of an irrational number is again this irrational number, and so in particular $\sqrt{2} = \langle 1, 2, 2, 2, 2, 2, \dots \rangle$.

17.60 Theorem. *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Then*

$$\alpha = \langle \lfloor \alpha \rfloor, \lfloor \varphi(\alpha) \rfloor, \lfloor \varphi^2(\alpha) \rfloor, \lfloor \varphi^3(\alpha) \rfloor, \dots \rangle.$$

Proof. We write $a_n = \lfloor \varphi^{n-1}(\alpha) \rfloor$ for $n \in \mathbb{N}^+$. To prove

$$\alpha = \lim_n \langle a_1, a_2, \dots, a_n \rangle.$$

This follows from

$$\begin{aligned} |\alpha - \langle a_1, \dots, a_n \rangle| &= |\langle a_1, \dots, a_n, \varphi^n(\alpha) \rangle - \langle a_1, \dots, a_n \rangle| \\ &= \frac{1}{|q_n(a_1, \dots, a_n) q_{n+1}(a_1, \dots, a_n, \varphi^n(\alpha))|} < \frac{1}{q_n(a_1, \dots, a_n)^2} \end{aligned}$$

and the fact that the numbers $q_n(a_1, \dots, a_n)$ from $n = 2$ onwards form a strict ascending sequence. □

17.61 Definition. The rational number $\langle \lfloor \alpha \rfloor, \lfloor \varphi(\alpha) \rfloor, \lfloor \varphi^2(\alpha) \rfloor, \dots, \lfloor \varphi^{n-1}(\alpha) \rfloor \rangle$ is called the n-th *convergent* of the irrational number α.

17.62 Lemma. *Let $\alpha = \langle a_1, a_2, a_3, \dots \rangle$ with $a_1 \in \mathbb{Z}$ and $a_2, a_3, \dots \in \mathbb{N}^+$. Then $\lfloor \alpha \rfloor = a_1$ and $\varphi(\alpha) = \langle a_2, a_3, \dots \rangle$.*

Proof.

$$\langle a_1, a_2, a_3, \dots \rangle = \lim_n \langle a_1, \dots, a_{n+1} \rangle = \lim_n a_1 + \frac{1}{\langle a_2, a_3, \dots, a_{n+1} \rangle}$$

Christiaan Huygens (The Hague 1629 – The Hague 1695)

Christiaan Huygens was the first to use continued fractions for concrete applications. He used them for the determination of the number of teeth of gear wheels to be used in a planetarium. He discovered the moon Titan of Saturn and also the ring around Saturn using a self-made lens. He was a friend of Descartes and corresponded with among others Mersenne, Pascal and Fermat. He wrote a book on probability calculus. He designed a pendulum clock for accurate time measurement. Huygens often stayed in Paris and London having contacts with among others Leibniz and Newton. He contributed to the foundations of mechanics and the theory of light.

$$= a_1 + \frac{1}{\lim_n \langle a_2, a_3, \ldots, a_{n+1} \rangle} = a_1 + \frac{1}{\langle a_2, a_3, \ldots \rangle}.$$

Since $a_2, a_3, \ldots \in \mathbb{N}^+$ we have $\langle a_2, a_3, \ldots \rangle > 1$, and so $\lfloor \alpha \rfloor = a_1$ and $\varphi(\alpha) = \langle a_2, a_3, \ldots \rangle$. □

17.63 Theorem. *Let $\alpha = \langle a_1, a_2, a_3, \ldots \rangle$ with $a_1 \in \mathbb{Z}$ and $a_2, a_3, \ldots \in \mathbb{N}^+$. Then for all $n \in \mathbb{N}^+$*

$$a_n = \lfloor \varphi^{n-1}(\alpha) \rfloor.$$

(So the sequence $a_1, a_2, a_3, \ldots$ is the continued fraction expansion of α.)

Proof. From lemma 17.62 it follows that for all $n \in \mathbb{N}^+$

$$\varphi^{n-1}(\alpha) = \langle a_n, a_{n+1}, \ldots \rangle,$$

and then by the same lemma $a_n = \lfloor \varphi^{n-1}(\alpha) \rfloor$. □

From the above it follows that the map from $\mathbb{R} \setminus \mathbb{Q}$ to the set of the sequences $a_1, a_2, a_3, \ldots$ with $a_1 \in \mathbb{Z}$ and $a_2, a_3 \ldots \in \mathbb{N}^+$, which maps an irrational number to its continued fraction expansion, is bijective. The inverse of this map assigns to such a sequence its continued fraction.

17.8 Diophantine Approximation

For a given irrational number α there is for every $n \in \mathbb{N}^+$ a unique $p \in \mathbb{Z}$ such that

$$\left| \alpha - \frac{p}{10^n} \right| < \frac{1}{2 \cdot 10^n}.$$

Thus α is approximated by the decimal fraction $\frac{p}{10^n}$, or in terms of the decimal notation, it is approximated up to n digits to the right of the decimal point. Diophantine approximation is about approximation by rationals and not just decimal approximation in which only fractions having a power of ten as denominator are used. For a given α some denominators are better suited for approximation than others. Given a denominator $q \in \mathbb{N}^+$, there is a unique numerator $p \in \mathbb{Z}$ such that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q}.$$

The number

$$|q\alpha - p|$$

is a good measure for the usefulness of the denominator q. We have

$$|q\alpha - p| < \tfrac{1}{2}.$$

We will show that the convergents of the continued fraction expansion do considerably better. Let the sequence $a_1, a_2, a_3, \dots$ be the continued fraction expansion of α (so $a_n = [\varphi^{n-1}(\alpha)]$ for $n = 1, 2, 3, \dots$). We write p_n for $p_n(a_1, \dots, a_n)$ and q_n for $q_n(a_1, \dots, a_n)$. We will see that the q_n are 'good' denominators for the approximation of α and that infinitely many of them are even 'very good' denominators.

17.64 Proposition. *$|q_n\alpha - p_n| < \frac{1}{q_n}$ for all $n \in \mathbb{N}^+$.*

Proof. Let $n \in \mathbb{N}^+$. We have $\frac{p_n}{q_n} < \alpha < \frac{p_{n+1}}{q_{n+1}}$ or $\frac{p_{n+1}}{q_{n+1}} < \alpha < \frac{p_n}{q_n}$, and so

$$\left|\alpha - \frac{p_n}{q_n}\right| < \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| = \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}.$$

Hence $|q_n\alpha - p_n| < \frac{1}{q_n}$. □

17.65 Proposition. *For infinitely many $n \in \mathbb{N}^+$ we have $|q_n\alpha - p_n| < \frac{1}{2q_n}$.*

Proof. For all $n \in \mathbb{N}^+$ we have

$$\left|\alpha - \frac{p_n}{q_n}\right| + \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right| = \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| = \frac{1}{q_n q_{n+1}} < \frac{1}{2q_n^2} + \frac{1}{2q_{n+1}^2}.$$

(Use: $ab < \frac{1}{2}(a^2 + b^2)$ if $a \neq b$.) So not both $|\alpha - \frac{p_n}{q_n}| \geq \frac{1}{2q_n^2}$ and $|\alpha - \frac{p_{n+1}}{q_{n+1}}| \geq \frac{1}{2q_{n+1}^2}$. Hence

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{2q_n^2} \qquad \text{or} \qquad \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right| < \frac{1}{2q_{n+1}^2}.$$ □

17.66 Lemma. *Let $n \in \mathbb{N}^+$. For all $q \in \mathbb{N}^+$ with $q < q_{n+1}$ and all $p \in \mathbb{Z}$ we have*

$$|q\alpha - p| \geq |q_n\alpha - p_n|.$$

Proof. There are $u, v \in \mathbb{Z}$ such that

$$\begin{aligned} p &= up_n + vp_{n+1} \\ q &= uq_n + vq_{n+1}. \end{aligned}$$

Multiply the first equation by q_n, the second by p_n and subtract. This yields $(-1)^n v = qp_n - pq_n$. Similarly, multiplication of the first by q_{n+1} and the second by p_{n+1} leads to $(-1)^n u = pq_{n+1} - qp_{n+1}$. If $u = 0$, then $q = vq_{n+1}$, contradictory to $q < q_{n+1}$. So $u \neq 0$. From $q = uq_n + vq_{n+1}$ it easily follows that u and v have different signs (if $v \neq 0$), and so (since this also holds for $q_n\alpha - p_n$ and $q_{n+1}\alpha - p_{n+1}$):

$$|q\alpha - p| = |u(q_n\alpha - p_n) + v(q_{n+1}\alpha - p_{n+1})| \geq |u(q_n\alpha - p_n)| \geq |q_n\alpha - p_n|. \quad \square$$

So for the approximation of α among the numbers with denominator $< q_{n+1}$ there is no better denominator than q_n.

17.67 Theorem. *Let $q \in \mathbb{N}^+$ and let $p \in \mathbb{Z}$ such that $|q\alpha - p| < \frac{1}{2q}$ and $\gcd(p, q) = 1$. Then there is an $n \in \mathbb{N}^+$ such that $p = p_n$ and $q = q_n$.*

Proof. There is a unique $n \in \mathbb{N}^+$ such that $q_n \leq q < q_{n+1}$. For this n we have by lemma 17.66:

$$|pq_n - p_n q| \leq |qq_n\alpha - pq_n| + |qq_n\alpha - p_n q| \leq (q_n + q)|q\alpha - p| < 2q \cdot \frac{1}{2q} = 1$$

Since $pq_n - p_n q$ is an integer, it must be equal to 0. So $pq_n - p_n q = 0$. Because $\gcd(p, q) = 1$ from this follows $q \mid q_n$, and since $\gcd(p_n, q_n) = 1$, also $q_n \mid q$. Hence $q = q_n$. $\square$

17.68 Example. In example 17.59 we computed the first 8 convergents of $\sqrt{2}$. The existence of $\sqrt{2}$ we already established using the supremum property. The convergents of $\sqrt{2}$ form a sequence of rational numbers converging to $\sqrt{2}$. The distance between $\sqrt{2}$ and the convergent $\frac{p_i}{q_i}$ is bounded by $\frac{1}{q_i q_{i+1}}$.

i	p_i	q_i	$\frac{p_i}{q_i}$	$\frac{p_i^2}{q_i^2}$	$\frac{1}{q_i q_{i+1}}$
1	1	1	1	1	0,5
2	3	2	1,5	2,25	0,1
3	7	5	1,4	1,96	0,01 ...
4	17	12	1,416666 ...	2,069444 ...	0,002 ...
5	41	29	1,413793 ...	1,998810 ...	0,0004 ...
6	99	70	1,414285 ...	2,000204 ...	0,00008 ...
7	239	169	1,414201 ...	1,999964 ...	0,00001 ...
8	577	408	1,414215 ...	2,000006 ...	0,000002 ...

For the number π we have

$$\pi = \langle 3, 7, 15, 1, 292, \dots \rangle.$$

And so

i:	-1	0	1	2	3	4	5	$\cdots$
a_i:	$-$	$-$	3	7	15	1	292	$\cdots$
p_i:	0	1	3	22	333	355	103993	$\cdots$
q_i:	1	0	1	7	106	113	33102	$\cdots$

Thus we have very good approximations of π:

$$\left|\pi - \frac{22}{7}\right| < \frac{1}{7 \cdot 106} = \frac{1}{742}.$$

Since a_5 is relatively large, $\frac{p_4}{q_4}$ is a particularly good approximation of π:

$$\left|\pi - \frac{355}{113}\right| < \frac{1}{113 \cdot 33102} < 3 \cdot 10^{-7}.$$

Nobody has been able to discover some kind of regularity in the continued fraction expansion of π. The number $\sqrt{2}$ does have such a regularity: the expansion repeats with a period of length 1. Other kinds of regularity are possible, e.g.

$$\langle 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, \dots \rangle.$$

Remarkably this is the case for the number e, the base of the natural logarithm:

$$e = \langle 2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1, 12, 1, 1, 14, \dots \rangle.$$

17.9 Uncountable Sets

A set is finite if it is equipotent to $\underline{n}$ for some natural number n. Sets which are equipotent to $\mathbb{N}$ are called countable. The sets $\mathbb{N} \times \mathbb{N}$, $\mathbb{Z}$, $\mathcal{R}_0(\mathbb{N})$ and $\mathbb{Q}$ are countable. See also the exercises 12, 13, 14 of chapter 5, exercise 11 of chapter 7, exercise 14 of chapter 8 and exercise 2 of chapter 9.

17.69 Definition. An infinite set which is not countable is called *uncountable*.

We will show the uncountability of some sets by a method of Cantor. By the same method it can be made clear that uncountable sets are not necessarily equipotent.

We start with the set $\mathcal{R}(\{0,1\})$, the set of sequences in $\{0,1\}$.

17.70 Theorem. *The set $\mathcal{R}(\{0,1\})$ is uncountable.*

Proof. Let be given a map $f\colon \mathbb{N} \to \mathcal{R}(\{0,1\})$. For every $m \in \mathbb{N}$ we have a sequence $f(m) = (f(m)_n)$. Consider the sequence (a_n) in $\{0,1\}$ defined by $a_n = 1 - f(n)_n$. Then in particular $a_n \neq f(n)_n$ for every $n \in \mathbb{N}$.

The sequence (a_n) differs from all sequences $f(m)$: let $m \in \mathbb{N}$, then the sequence (a_n) differs from the sequence $f(m)$, because the m-th term of (a_n) is $1 - f(m)_m$, the m-th term of $f(m)$ being $f(m)_m$.

So there is no surjective map from $\mathbb{N}$ to $\mathcal{R}(\{0,1\})$, let alone a bijective one. □

We make it more concrete (and less exact). An expression like

$$10100010111001001000111100 01\ldots$$

will stand for (the start of) a sequence in $\{0,1\}$ and suppose we have such a sequence for every $n \in \mathbb{N}$, say it starts as follows:

$$\begin{aligned}
&10100010111001001000111100 01\ldots\\
&01011110010100011111111000 101\ldots\\
&10011010101010010100000111 11\ldots\\
&01010110101010100001011010 11\ldots\\
&01010101001001010101010100 \ldots\\
&10010101010100010100101000 01\ldots\\
&11110000000000000000111111 111\ldots\\
&10101000000111111111100010 10\ldots\\
&00000011000000001100000001 00\ldots\\
&11001110010010101110010100 11\ldots\\
&\vdots
\end{aligned}$$

From this we can extract the *diagonal*, the sequence having as n-th term the n-th term of the n-th sequence:

$$1101011110\ldots$$

In this sequence we replace every 1 by a 0 and every 0 by a 1:

$$0010100001\ldots$$

This sequence does not occur in the sequence of sequences: for every n it differs from n-th sequence, since their nth terms are different.

This type of reasoning is known as Cantor's *diagonal argument*.

17.71 Corollary. *The set $\mathbb{R}$ is uncountable.*

Proof. The map $\mathcal{R}(\{0,1\}) \to \mathbb{R},\ (c_n) \mapsto \sum_{n=0}^{\infty} \frac{c_n}{10^n}$ is injective. So $\mathbb{R}$ contains an uncountable subset and therefore it cannot be countable. □

We use the notation $I(a,b) = \{\, x \in \mathbb{R} \mid a < x < b\}$ for $a, b \in \mathbb{R}$ with $a < b$. The proof of Corollary 17.71 shows that for example the subset $I(-1,1)$ is uncountable. In fact this set is equipotent to $\mathbb{R}$:

17.72 Proposition. $\{\, x \in \mathbb{R} \mid -1 < x < 1\} \approx \mathbb{R}$

Proof. We write $I = \{\, x \in \mathbb{R} \mid -1 < x < 1\}$ and define a map $f\colon\ I \to \mathbb{R}$ by

$$f(x) = \frac{x}{1-x^2} \qquad (\text{for all } x \in I).$$

We will show that f is bijective. Let $a \in \mathbb{R}$. Is there a unique $x \in I$ with $f(x) = a$? Such an x is the solution of the quadratic equation $ax^2 + x - a = 0$. The discriminant is $1 + 4a^2 > 0$. The solutions are $x = \frac{-1 \pm \sqrt{1+4a^2}}{2a}$. Exactly one of these is an element of I. □

The rational numbers form a countable subset of $\mathbb{R}$. As a consequence the irrational numbers form an uncountable subset of $\mathbb{R}$: if it was countable, then $\mathbb{R}$, being the union of two countable sets, would be countable as well. But what about the subset of algebraic numbers?

17.73 Proposition. *The subset of $\mathbb{R}$ consisting of algebraic numbers is countable.*

Proof. An algebraic number is a zero of a polynomial $\neq 0$ having coefficients in $\mathbb{Q}$. Multiplication by a multiple of the denominators of the coefficients gives a polynomial having coefficients in $\mathbb{Z}$. So algebraic numbers are zeros of such polynomials. These polynomials correspond to elements of $\mathcal{R}_0(\mathbb{Z})$ and thus form a countable set. Since each of these polynomials has only a finite number of zeros, the set of all zeros of these polynomials is countable as well. □

17.74 Corollary. *The subset of $\mathbb{R}$ consisting of the transcendental numbers is uncountable.* □

We will order sets by magnitude.

17.75 Definition. Let A and B be sets. We define

$$A \preceq B \iff \text{there is an injective map from } A \text{ to } B.$$

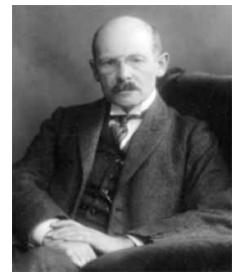
and

$$A \prec B \iff A \preceq B \text{ and there is no surjective map } A \to B.$$

The following theorem of Cantor, Schröder and Bernstein allows us in many cases to see that sets are equipotent.

Friedrich Wilhelm Karl Ernst Schröder (Mannheim 1841 – Karlsruhe 1902)
Felix Bernstein (Halle 1878 – Zürich 1956)

The German mathematicians Ernst Schröder (left) and Felix Bernstein (right) proved theorem 17.76, which was first published by Cantor, however without giving a proof.



17.76 Theorem (Cantor, Schröder, Bernstein). *Let A and B be sets with $A \preceq B$ and $B \preceq A$. Then $A \approx B$.*

Proof. We assume that $A \cap B = \emptyset$. This is allowed because we may replace A by $A \times \{0\}$ and B by $B \times \{1\}$. Let $f\colon A \to B$ and $g\colon B \to A$ be injective maps. We define a transformation F of $A \cup B$:

$$F(x) = \begin{cases} f(x) & \text{if } x \in A \\ g(x) & \text{if } x \in B. \end{cases}$$

This transformation F is injective: if $F(x) = F(y)$, then $x, y \in A$ or $x, y \in B$ and so $x = y$, because f and g are injective. Call $x \in A \cup B$ a *successor* of y if $F(y) = x$. We then also call y a *predecessor* of x. So for every $x \in A \cup B$ there is a unique successor and at most one predecessor since F is injective. Let A_0 be the subset of A of elements having no predecessor and B_0 the subset of B of elements having no predecessor. We define a map $h\colon A \to B$:

$$h(a) = \begin{cases} \text{the predecessor of } a, & \text{if } a \text{ is in the course of an element of } B_0; \\ \text{the successor of } a, & \text{otherwise.} \end{cases}$$

This h is bijective. The inverse is $k\colon B \to A$ defined by:

$$k(b) = \begin{cases} \text{the successor of } b, & \text{if } b \text{ is in the course of an element of } B_0; \\ \text{the successor of } b, & \text{otherwise.} \end{cases}$$

□

17.77 Examples.

a) $\{\, x \in \mathbb{R} \mid 0 \le x \le 1 \,\} \approx I(0,1)$. An injective map from $\{\, x \in \mathbb{R} \mid 0 \le x \le 1 \,\}$ to I is for example $x \mapsto \frac{1}{2}x + \frac{1}{4}$.

b) It is seen directly that $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$ by giving a bijection. It also follows from the existence of injections $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$ and $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$:

$$\mathbb{N} \to \mathbb{N} \times \mathbb{N},\ n \mapsto (n, 0) \qquad \text{and} \qquad \mathbb{N} \times \mathbb{N} \to \mathbb{N},\ (m, n) \mapsto 2^m 3^n.$$

c) From corollary 17.71 and proposition 17.72 it follows that

$$\mathcal{R}(\{0,1\}) \preceq I(-1,1) \approx \mathbb{R}.$$

The map $I(0,1) \to \mathcal{R}(\{0,1\})$ that assigns to a real number its binary expansion, is injective. Also $I(0,1) \approx I(-1,1)$: map x to $2x - 1$. So:

$$I(-1,1) \approx I(0,1) \preceq \mathcal{R}(\{0,1\}) \preceq I(-1,1) \approx \mathbb{R}.$$

So these sets all are equipotent.

d) We show that $\mathbb{R} \times \mathbb{R} \approx \mathbb{R}$. Because $\mathbb{R} \approx \mathcal{R}(\{0,1\})$, we can show as well that $\mathcal{R}(\{0,1\}) \times \mathcal{R}(\{0,1\}) \approx \mathcal{R}(\{0,1\})$. The following map is a bijection:

$$((a_n), (b_n)) \mapsto (c_n) \qquad \text{with} \quad c_n = \begin{cases} a_{\frac{n}{2}} & \text{if } n \text{ even} \\ b_{\frac{n+1}{2}} & \text{if } n \text{ odd.} \end{cases}$$

We have $\mathbb{N} \prec \mathcal{R}(\{0,1\})$ and $\mathbb{N} \prec \mathbb{R}$. The set $\mathcal{R}(\{0,1\}) = \{0,1\}^{\mathbb{N}}$ is equipotent to $\mathcal{P}(\mathbb{N})$, see proposition 5.46. So we have $\mathbb{N} \prec \mathcal{P}(\mathbb{N})$. The method of Cantor is easily generalized to $A \prec \mathcal{P}(A)$:

17.78 Theorem. *Let A be a set. Then $A \prec \mathcal{P}(A)$.*

Proof. The map $A \to \mathcal{P}(A),\ a \mapsto \{a\}$ is injective. Let $f\colon A \to \mathcal{P}(A)$ be a map. We show that f is not surjective by showing that the set $U = \{\, x \in A \mid x \notin f(x) \,\}$ is not an image of an element of A.

Suppose $U = f(a)$ for an $a \in A$. We look at the element a.

Suppose $a \in U$. Then $a \notin f(a)$ and so $a \notin U$, since $U = f(a)$. Contradiction.

So $a \notin U$. But then not $a \notin f(a)$, that is $a \in f(a)$. So $a \in U$, since $f(a) = U$. Contradiction.

Hence there is no $a \in A$ with $U = f(a)$. So f is not surjective. □

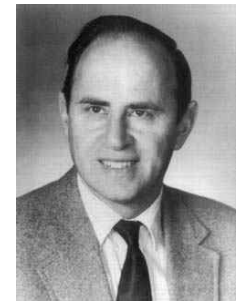
So $\mathbb{N} \prec \mathcal{P}(\mathbb{N}) \prec \mathcal{P}(\mathcal{P}(\mathbb{N})) \prec \mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N}))) \prec \cdots$. These sets are mutually not equipotent as follows directly from:

17.79 Lemma. *Let A, B and C be sets with $A \prec B \prec C$. Then $A \prec C$.*

Proof. Clearly there is an injective map $A \to C$. Suppose there is also a surjective one, say $f\colon A \to C$. Then there is a surjective one from B to C: just extend f to a map $B \to C$. Contradiction. □

Kurt Gödel (Brünn (now Brno) 1906 – Princeton (New Jersey) 1978)
Paul Cohen (Long Branch (New Jersey) 1934 – Palo Alto (California) 2007)

The Austrian logician Kurt Gödel (left) proved that in an axiom system for set theory containing Peano's axioms there are true propositions which cannot be proved nor disproved from the axioms. This is known as Gödel's *Incompleteness Theorem*. The American mathematician Paul Cohen (right) developed a way, known as *forcing*, of constructing mathematical models to test a hypothesis.



The standard finite sets $\underline{n}$ are used to indicate the number of elements of a finite set: $\#(A) = n$ if $A \approx \underline{n}$. For the countable sets $\mathbb{N}$ can be used as a standard set. It goes with a new 'number': $\aleph_0$. Hence: $\#(A) = \aleph_0$ if $A \approx \mathbb{N}$. ($\aleph$ is aleph, the first letter of the Hebrew alphabet.)

There are ways in set theory to define the standard sets, but here we will not dwell on this. These standard sets correspond to so-called *cardinal numbers*. Cardinal numbers can be ordered using $\preceq$: $\#(A) \leq \#(B) \iff A \preceq B$. Reflexivity and transitivity of $\leq$ are clear. The antisymmetry follows from the theorem of Cantor, Schröder and Bernstein. It can be shown using the axiom of choice that $A \preceq B$ of $B \preceq A$ for any pair of sets A, B. It can also be shown that for every cardinal number there is a least cardinal number which is greater, the *successor* of the cardinal number. The successor of $\aleph_0$ is denoted by $\aleph_1$, etc.

Cantor denoted the cardinal number of $\mathbb{R}$ by $\mathfrak{c}$. Is $\mathfrak{c} = \aleph_1$? That is: is there a cardinal number greater than $\aleph_0$ and less than $\mathfrak{c}$? The *continuum hypothesis* says that such a cardinal number does not exist. In 1940 **Kurt Gödel** showed that with the usual axioms for set theory the continuum hypothesis can not be falsified. In 1963 **Paul Cohen** proved that it can not be derived from axioms either.

For cardinal numbers operations can be defined using operations with sets:

Addition: $\#(A) + \#(B) = \#(A \cup B)$ if A and B are disjoint. The addition is associative and commutative. If one of the cardinal numbers is infinite, then the sum is the greatest of the two.

Multiplication: $\#(A) \cdot \#(B) = \#(A \times B)$. The multiplication is associative and commutative. It also is distributive over addition. If one of the cardinal numbers is infinite and if both are not 0, then the product is the greatest of the two.

Exponentiation: $\#(A)^{\#(B)} = \#(A^B)$. The usual rules hold here as well: $a^{b+c} = a^b a^c$, $a^{bc} = (a^b)^c$, $(ab)^c = a^c b^c$.

Take care with cancellation laws, opposites and inverses. For example we have seen that:

- $\aleph_0 \cdot \aleph_0 = \aleph_0$, because $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$.
- $2^{\aleph_0} = \mathfrak{c}$, because $\{0,1\}^{\mathbb{N}} \approx \mathbb{R}$.
- $\mathfrak{c} \cdot \mathfrak{c} = \mathfrak{c}$, because $\mathbb{R} \times \mathbb{R} \approx \mathbb{R}$.

Exercises

1. Let α be a real number > 1. Take $a \in \mathbb{N}$ such that $(a-1)^2 < \alpha \leq a^2$. The sequence (a_n) of real numbers is defined by

$$\begin{cases} a_0 = a \\ a_{n+1} = \frac{1}{2}(a_n + \frac{\alpha}{a_n}) \quad \text{for all } n \in \mathbb{N} \end{cases}$$

   (i) Show that $a_n^2 \geq \alpha$ for all $n \in \mathbb{N}$.

   (ii) Prove that the sequence (a_n) is descending.

   (iii) Show that $\lim_n a_n = \sqrt{\alpha}$.

   (This method for the approximation $\sqrt{\alpha}$ rests on Newton's method for the approximation of a zero of a function, in this case the function $x^2 - \alpha$.)

2. The sequence of rational numbers $(a_n)_{n\geq 1}$ is defined by $a_n = \sum_{k=1}^{n} \frac{1}{k(k+1)}$.

   (i) Show that $a_n = 1 - \frac{1}{n+1}$ for all $n \in \mathbb{N}^+$.

   (ii) Verify that $\frac{1}{n(n+1)} < \frac{1}{n^2} < \frac{1}{(n-1)n}$ for $n = 2, 3, 4, \ldots$.

   (iii) Show that $\frac{3}{2} < \sum_{n=1}^{\infty} \frac{1}{n^2} < 2$.

3. Let $\alpha \in \mathbb{R}$. Let the sequence $(a_n)_{n\geq 1}$ in $\mathbb{Z}$ be defined by

$$a_n = \lfloor n\alpha \rfloor \quad \text{for all } n \in \mathbb{N}.$$

   Show that conversely α is determined by this sequence. (In terms of maps: the map $\alpha \mapsto (a_n)$ from $\mathbb{R}$ to the set of sequences in $\mathbb{Z}$ is injective.)

4. Let $U \subseteq \mathbb{R}$ and let $f, g \colon U \to \mathbb{R}$ be continuous. Show that the functions

$$U \to \mathbb{R},\ x \mapsto f(x) + g(x),$$
$$U \to \mathbb{R},\ x \mapsto f(x)g(x)$$

   are continuous.

5. Let $g \colon \mathbb{R} \to \mathbb{R}$ be continuous with $g(x) \neq 0$ for all $x \in \mathbb{R}$. Given is that $g(0) > 0$. Show that $g(x) > 0$ for all $x \in \mathbb{R}$.

6. Let $U \subseteq \mathbb{R}$ and let $g \colon U \to \mathbb{R}$ be continuous with $g(x) \neq 0$ for all $x \in U$. Prove that the function $U \to \mathbb{R},\ x \mapsto \frac{1}{g(x)}$ is continuous.

7. Let U and V be subsets of $\mathbb{R}$ and let $f\colon U \to \mathbb{R}$ and $g\colon V \to \mathbb{R}$ be continuous functions. Let be given that $f_*(U) \subseteq V$. Prove that the function

$$U \to \mathbb{R},\ x \mapsto g(f(x))$$

is continuous.

8. Let $f(x)$ and $g(x)$ be polynomials with $g(x)$ not the 0-polynomial. Prove that the function

$$U \to \mathbb{R},\ x \mapsto \frac{f(x)}{g(x)},$$

where $U = \{\, x \in \mathbb{R} \mid g(x) \neq 0 \,\}$, is continuous.

9. Show that the map $\mathbb{R} \to \mathbb{R}^+,\ x \mapsto 2^x$ is an isomorphism of groups.

10. Let $\beta \in \mathbb{R}$ with $\beta > 1$. Let x and y be real numbers with $x < y$. Show that $\beta^x < \beta^y$.

11. Determine the continued fraction expansions of $\sqrt{13}$ and $\frac{\sqrt{13}+1}{2}$.

12. The irrational number $\alpha = \langle 1, 3, 1, 3, 1, 3, 1, 3, \ldots \rangle$ satisfies $\alpha = \langle 1, 3, \alpha \rangle$. Deduce from this that $\alpha = \frac{\sqrt{21}+3}{6}$.

13. Let $m \in \mathbb{N}^+$. Determine the continued fraction expansion of $\frac{m+\sqrt{m^2+4}}{2}$.

14. Let $0{,}b_1b_2b_3\ldots$ be the decimal notation of an irrational $\alpha \in [0, 1)$. We consider the continued fraction expansion of α. Suppose that there is an $n \in \mathbb{N}^+$ such that $q_n = 100$. Prove that $b_3 = b_4 = 0$ or $b_3 = b_4 = 9$.

15. Approximate $\sqrt{3}$ as well as possible with a rational number $\frac{p}{q}$, where $0 < q < 100$.

16. The continued fraction expansion gives a bijective map from the set of irrational numbers > 1 to $\mathcal{R}(\mathbb{N}^+)$. Verify this.

17. Give a bijective map from $\mathcal{R}(\mathbb{N})$ to $\mathcal{R}(\{0, 1\})$.

18. Let A be the set of sequences $(a_n)_{n\geq 0}$ with the property that $a_n \neq a_{n+1}$ for all $n \in \mathbb{N}$. Show that $A \approx \mathbb{R}$.

19. Does it hold for every convergent sequence (a_n) in $\mathbb{R}$, with $a_n \neq 0$ for all n, that

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 1\ ?$$

Give a proof or a counterexample.

20. (i) Give an example of a sequence (a_n) in $\mathbb{R}$ for which the following holds: (a_n) converges while the sequence of the floors $(\lfloor a_n \rfloor)$ does not.

    (ii) Let (a_n) be a convergent sequence in $\mathbb{R}$. Show that there are an $m \in \mathbb{Z}$ and a $N \in \mathbb{N}^+$ such that $m - 1 < a_n < m + 1$ for all $n \geq N$.

    (iii) Let (a_n) be a Cauchy sequence in $\mathbb{Q}$ with a diverging sequence $(\lfloor a_n \rfloor)$ of floors. Prove that (a_n) converges to an integer.

21. The real number α is given as an infinite continued fraction:

$$\alpha = \langle \overline{3} \rangle \ (= \langle 3, 3, 3, \dots \rangle).$$

(i) Show that $\alpha = \frac{3+\sqrt{13}}{2}$.

(ii) Determine natural numbers p and q such that

$$q < 100 \quad \text{and} \quad |\alpha - \tfrac{p}{q}| < \tfrac{1}{1000}.$$

22. Verify for each of the following assertions its truth. Give a proof or a counterexample.

(i) If α is an irrational number with $\alpha > 0$, then $\sqrt{\alpha}$ is an irrational number.

(ii) If α is an irrational number with $\alpha > 1$, then $\sqrt{\frac{\alpha+1}{\alpha-1}}$ is an irrational number.

(iii) If α is an irrational number, then $\sqrt{\alpha^2+1}$ is an irrational number.

23. On the set $\mathcal{R}(\{0,1\})$ (consisting of all infinite sequences of ones and zeros) the relation $\sim$ is defined by

$$(a_0, a_1, a_2, \dots) \sim (b_0, b_1, b_2, \dots) \iff a_n \neq b_n \text{ for only finitely many } n \in \mathbb{N}.$$

(i) Prove that $\sim$ is an equivalence relation.

(ii) Show that every equivalence class is countable.

(iii) Prove that there are uncountably many equivalence classes.

24. (i) Let B be the set of all sequences $a_0, a_1, a_2, \dots$ in $\{0,1,2\}$ with the property

$$a_{n+1} \neq a_n \text{ for all } n \in \mathbb{N}.$$

Is B finite, is B countable or is B uncountable?

(ii) Let C be the set of all sequences $a_0, a_1, a_2, \dots$ in $\{0,1,2\}$ with the property

$$\#(\{a_n, a_{n+1}, a_{n+2}\}) \neq 2 \text{ for all } n \in \mathbb{N}.$$

Is C finite, is C countable or is C uncountable?

25. (i) Let $g\colon \mathbb{N} \to \mathbb{N}$. Prove that

$$g \text{ is injective} \iff g(n+1) \notin \{g(0), g(1), \dots, g(n)\} \text{ for all } n \in \mathbb{N}.$$

(ii) Prove that the set $\{\, f\colon \mathbb{N} \to \mathbb{N} \mid f \text{ is injective} \,\}$ is uncountable.

26. For every sequence $(a_1, a_2, a_3, \dots)$ in $\{1,2,3\}$ we have real numbers

$$x = \sum_{n=1}^{\infty} \frac{x_n}{2^n}, \quad y = \sum_{n=1}^{\infty} \frac{y_n}{2^n} \quad \text{and} \quad z = \sum_{n=1}^{\infty} \frac{z_n}{2^n},$$

where

$$x_n = \begin{cases} 1 & \text{if } a_n = 1 \\ 0 & \text{otherwise,} \end{cases} \quad y_n = \begin{cases} 1 & \text{if } a_n = 2 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad z_n = \begin{cases} 1 & \text{if } a_n = 3 \\ 0 & \text{otherwise.} \end{cases}$$

(i) Let $(a_1, a_2, a_3, \dots)$ be a sequence in $\{1, 2, 3\}$. Show that $x + y + z = 1$.

(ii) Let be given that the sequence (a_n) repeats. Show that $x, y, z \in \mathbb{Q}$.

(iii) Let (a_n) be a sequence in $\{1, 2, 3\}$ with $x \in \mathbb{Q}$. Does it follow that the sequence (a_n) repeats?

(iv) Let (a_n) be the sequence $(1, 2, 3, 1, 2, 3, 1, 2, 3, \dots) = (\overline{1, 2, 3})$. Compute for this sequence the numbers x, y and z. Write them as an ordinary fraction.

27. Let A be the set of all arithmetic progressions of real numbers and let B be the set of all sequences $r_0, r_1, r_2, \dots$ of real numbers with $r_{n+2} = 2r_{n+1} - r_n$ for all $n \in \mathbb{N}$. Prove that $A = B$.

28. The sequence (a_n) in $\mathbb{Q}$ is given by

$$a_n = \sum_{k=0}^{n} \frac{(-1)^k}{2^k} \qquad \text{(for all } n \in \mathbb{N}\text{)}.$$

(i) Show that the sequence (a_n) converges in $\mathbb{Q}$.

(ii) Give an $N \in \mathbb{N}$ such that $|a_n - \lim_{n\to\infty} a_n| < 10^{-100}$ for all $n \geq N$.

29. Verify for each of the following real numbers whether they are rational:

$$\sqrt[3]{189}, \quad \sqrt[3]{\frac{189}{56}}, \quad \sqrt{3 + 2\sqrt{2}}, \quad \frac{\sqrt{6 + 2\sqrt{5}}}{1 + \sqrt{5}}.$$

30. The sequence $a_0, a_1, a_2, \dots$ of natural numbers is defined by

$$\begin{cases} a_0 = 4, \\ a_{n+1} = a_n^2 - 2 \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

Prove that

$$a_n = (2 + \sqrt{3})^{2^n} + (2 - \sqrt{3})^{2^n}$$

for all $n \in \mathbb{N}$.

31. Let $\beta \in \mathbb{R}$ with $\beta \leq 1$.

(i) Show that $n^\beta > n$ for all $n \in \mathbb{N}^+$.

(ii) Prove that the series $\sum_{n=1}^{\infty} \frac{1}{n^\beta}$ diverges.

32. Let $\beta \in \mathbb{R}$ with $\beta > 1$. The sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ are given by

$$a_n = \sum_{k=1}^{n} \frac{1}{k^\beta} \quad \text{and} \quad b_n = a_n + \frac{1}{n^{\beta-1}}.$$

(i) Show that the sequence (b_n) is descending.

(ii) Show that the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^\beta}$ converges.

18 The p-Adic Numbers

In this chapter p is a fixed arbitrary prime number. The examples will be about concrete prime numbers. In chapter 17 we completed $\mathbb{Q}$ with respect to the ordinary absolute value. Completion with respect to the p-adic absolute value yields the field $\mathbb{Q}_p$ of the p-adic numbers. In section 18.1 this construction is made. Also for this field we will have a close look at the multiplicative group. Via exponential functions this group too is closely related to the additive group. The field $\mathbb{Q}_p$ is not as common as $\mathbb{R}$ is. Outside mathematics there are hardly any applications. Inside mathematics the main applications are in number theory. In chapter 20 an example of such an application is given.

18.1 Construction of $\mathbb{Q}_p$

In this section $\mathbb{Q}$ is completed with respect to the p-adic absolute value. Here we follow the same route as we did in the previous chapter for the ordinary absolute value and do so with special attention to some remarkable properties.

18.1.1 The set $\mathbb{Q}_p$

18.1 Definition. p-Adic Cauchy sequences (a_n) and (b_n) in $\mathbb{Q}$ are called *p-adically equivalent* if the sequence $(a_n - b_n)$ is a p-adic null sequence. Notation: $(a_n) \sim_p (b_n)$. We denote the set of p-adic Cauchy sequences in $\mathbb{Q}$ by $\mathrm{CS}_p(\mathbb{Q})$.

Also in this case we have:

18.2 Proposition. *The relation $\sim_p$ in $\mathrm{CS}_p(\mathbb{Q})$ is an equivalence relation.* □

18.3 Definition. A *p-adic number* is an equivalence class in $\mathrm{CS}_p(\mathbb{Q})$ for the relation $\sim_p$. Notation: the class of a p-adic Cauchy sequence (a_n) will be denoted by $[(a_n)]$. The set $\mathbb{Q}_p$ is the set of the p-adic numbers.

In this chapter we consider only the p-adic case. If we write $\alpha = [(a_n)]$, then we always mean that (a_n) is a p-adic Cauchy sequence and α is the equivalence class with respect to the relation $\sim_p$, thus being a p-adic number.

18.1.2 The field $\mathbb{Q}_p$

The sum and the product of p-adic numbers is obtained by adding and multiplying representing Cauchy sequences. This is independent of the choice of representatives:

18.4 Definition. Let (a_n) and (b_n) be p-adic Cauchy sequences in $\mathbb{Q}$. The *sum* and the *product* of the p-adic numbers $[(a_n)]$ and $[(b_n)]$ are defined by

$$[(a_n)] + [(b_n)] = [(a_n + b_n)]$$
$$[(a_n)] \cdot [(b_n)] = [(a_n b_n)].$$

Again we have an injective map

$$\mathbb{Q} \to \mathbb{Q}_p,\ a \mapsto [(a)].$$

We can see $\mathbb{Q}_p$ as an extension of $\mathbb{Q}$. The class $[(a)]$ of a constant sequence (a) will usually be denoted by a. It is the class of sequences in $\mathbb{Q}$ converging p-adically to a.

18.5 Theorem. *The set $\mathbb{Q}_p$ together with the addition and the multiplication is a field.*

Proof. We only look at the existence of inverses. Let $\alpha \in \mathbb{Q}_p$ with $\alpha \neq 0$. Choose a representative of α. Then (a_n) is not a p-adic null sequence. From proposition 16.81 it follows that there is an $N \in \mathbb{N}$ such that $|a_n|_p = |a_N|_p$ for all $n \geq N$. So we can assume that $a_n \neq 0$ for all n. From proposition 16.84 it then follows that the sequence $(\frac{1}{a_n})$ is a p-adic Cauchy sequence as well. We have $[(a_n)][(\frac{1}{a_n})] = 1$. □

18.1.3 The absolute value on $\mathbb{Q}_p$

We extend the p-adic absolute value on $\mathbb{Q}$ to an absolute value on $\mathbb{Q}_p$. If α is a p-adic number $\neq 0$, say $\alpha = [(a_n)]$, then there is an $N \in \mathbb{N}$ with $|a_n|_p = |a_N|_p$ for all $n \geq N$.

18.6 Definition. Let α be a p-adic number. Then we define the *absolute value* $|\alpha|_p$ as follows

$$|\alpha|_p = \begin{cases} |a_N|_p & \text{if } \alpha \neq 0 \text{ (N being as above),} \\ 0 & \text{if } \alpha = 0. \end{cases}$$

So we have $|\alpha|_p = \lim_n |a_n|_p$, the limit being the ordinary limit in $\mathbb{Q}$. If $\alpha \neq 0$, the sequence $|a_n|$ is constant for large n. If $\alpha = 0$, then the sequence $|a_n|_p$ is a null sequence. The absolute values of p-adic numbers are numbers in $\mathbb{R}^{\geq 0}$. The image of $\alpha \mapsto |\alpha|_p$ is the set $\{\, p^n \mid n \in \mathbb{Z} \,\} \cup \{0\}$.

From the definition of the absolute value on $\mathbb{Q}_p$ immediately follows:

18.7 Proposition. *The absolute value $\mathbb{Q}_p \to \mathbb{R}^{\geq 0}$, $\alpha \mapsto |\alpha|_p$ is a non-Archimedean absolute value: for all $\alpha, \beta \in \mathbb{Q}_p$ we have*

(i) $|\alpha|_p = 0 \iff \alpha = 0$,
(ii) $|\alpha\beta|_p = |\alpha|_p|\beta|_p$,
(iii) $|\alpha + \beta|_p \leq \max(|\alpha|_p, |\beta|_p)$. □

18.2 The Completeness of $\mathbb{Q}_p$

The p-adic absolute value on $\mathbb{Q}$ has been extended to $\mathbb{Q}_p$. Thus we now have null sequences, convergent sequences and Cauchy sequences in $\mathbb{Q}_p$. We will first show that every p-adic Cauchy sequence in $\mathbb{Q}$ converges in $\mathbb{Q}_p$.

18.8 Proposition. *Let (a_n) be a p-adic Cauchy sequence in $\mathbb{Q}$. Then it converges in $\mathbb{Q}_p$ to $\alpha = [(a_n)]$.*

Proof. Let $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$. Then there is an $N \in \mathbb{N}$ such that $|a_m - a_n|_p < \varepsilon$ for all $m, n > N$. For fixed m the sequence is $(a_m - a_n)$ a p-adic Cauchy sequence and we have $[(a_m - a_n)] = a_m - \alpha$. For $m \geq N$ we have that $|a_m - a_n|_p < \varepsilon$ for all $n \geq N$. So $|a_m - \alpha|_p = \lim_n |a_m - a_n|_p \leq \varepsilon$ for all $m \geq N$, that is (a_n) converges in $\mathbb{Q}_p$ to α. □

So near every p-adic number there is a rational number within any prescribed distance:

18.9 Corollary. *Let α be a p-adic number and let $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$. Then there is a rational number a with $|a - \alpha|_p < \varepsilon$.* □

It follows that $\mathbb{Q}_p$ is complete:

18.10 Theorem. *Let (α_n) be a Cauchy sequence of p-adic numbers. Then (α_n) converges in $\mathbb{Q}_p$.*

Proof. Analogous to the proof of theorem 17.18. □

We have used properties of null sequences, convergent sequences and Cauchy sequences in $\mathbb{Q}_p$ which were already proved for p-adic Cauchy sequences in $\mathbb{Q}$. All notions and properties of section 16.6 are applicable to the completion $\mathbb{Q}_p$. In particular we have:

18.11 Theorem. *A sequence in $\mathbb{Q}_p$ converges if and only if its difference sequence is a null sequence.* □

So for *any* sequence $(c_n) \in \mathbb{N}_p$ the sequence $(\sum_{k=0}^{n} c_k p^k)$ in $\mathbb{Q}_p$ converges, see Corollary 16.79. We will see that conversely every $\alpha \in \mathbb{Q}_p$ with $|\alpha|_p \leq 1$ has a p-adic expansion which does not necessarily repeat.

18.3 The Ring $\mathbb{Z}_p$

18.12 Definition. A p-adic number α is called *integral* or a p-adic *integer* if $|\alpha|_p \leq 1$. The set of integral p-adic numbers is denoted by $\mathbb{Z}_p$.

> In chapter 16 we considered the ring $\mathbb{Z}_{(p)}$. Also $\mathbb{Z}_p$ is a ring and we have $\mathbb{Z}_{(p)} = \mathbb{Z}_p \cap \mathbb{Q}$. The group $\mathbb{Z}_p^*$ consists of the p-adic numbers α with $|\alpha|_p = 1$. Recall that $\mathbb{N}_p$ is the set $\{0, \ldots, p-1\}$ (Notations 5.29).

18.13 Lemma. *Let $\alpha \in \mathbb{Z}_p$. Then there is a unique $c \in \mathbb{N}_p$ with $|\alpha - c|_p < 1$.*

Proof. Take an $r \in \mathbb{Q}$ with $|\alpha - r|_p < 1$. Then $|r|_p = |r - \alpha + \alpha|_p \leq \max(|r - \alpha|_p, |\alpha|_p) \leq 1$. So $r \in \mathbb{Z}_{(p)}$. By proposition 16.71 there exists a $c \in \mathbb{N}_p$ with $|r - c|_p < 1$. For this c we have $|\alpha - c|_p = |\alpha - r + r - c|_p \leq \max(|\alpha - r|_p, |r - c|_p) < 1$. □

18.14 Definition. Let $\alpha \in \mathbb{Z}_p$. The unique $c \in \mathbb{N}_p$ with $|\alpha - c|_p < 1$ is called the *remainder* of α after division by p. Notation: $c = [\alpha]_p$. (Thus we have extended division with remainder in $\mathbb{Z}_{(p)}$ to $\mathbb{Z}_p$.)

Also the transformation γ_p of $\mathbb{Z}_{(p)}$ as defined in chapter 16 can be extended to a transformation of $\mathbb{Z}_p$:

$$\gamma_p \colon \mathbb{Z}_p \to \mathbb{Z}_p,\ \alpha \mapsto \frac{\alpha - [\alpha]_p}{p}.$$

It can be used for p-adic expansions of p-adic integers in general:

18.15 Definition. Let $\alpha \in \mathbb{Z}_p$. The sequence $(c_n)_{n \geq 0}$ with $c_n = [\gamma_p^n(\alpha)]_p$ in $\mathbb{N}_p$ is called the *p-adic expansion* of α.

Now we have a bijection

$$\mathcal{R}(\mathbb{N}_p) \to \mathbb{Z}_p, \quad (c_n) \mapsto \sum_{n=0}^{\infty} c_n p^n$$

and its inverse

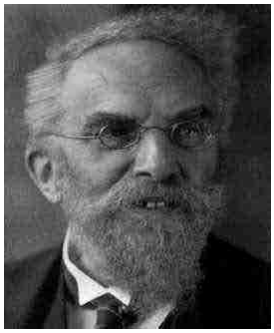
$$\mathbb{Z}_p \to \mathcal{R}(\mathbb{N}_p), \quad \alpha \mapsto ([\gamma_p^n(\alpha)]_p).$$

The repeating sequences correspond with the elements of $\mathbb{Z}_{(p)}$.

18.3.1 Another description of $\mathbb{Z}_p$

As we have seen, p-adic integers have a unique p-adic expansion. Therefore, the ring $\mathbb{Z}_p$ could have been constructed differently, namely as the set $\mathcal{R}(\mathbb{N}_p)$, where (c_n) is to be interpreted as $\sum_{n=0}^{\infty} c_n p^n$. In principle $\mathbb{R}$ too could have been constructed using the decimal notation of real numbers. The big problem in that case is the definition of addition and multiplication: for performing these operations one works from right to left, but there is no right end where to start. Moreover, in such an

Kurt Hensel (Königsberg (now Kaliningrad) 1861 – Marburg 1941)



Hensel was the first to study p-adic numbers. He introduced them as formal sums $\sum_{n=m}^{\infty} c_n p^n$ with $m \in \mathbb{Z}$ (for $m = 0$ these represent the p-adic integers). He used them for representing numbers by rational quadratic forms, e.g. forms of type $x^2 - ay^2$ with $a \in \mathbb{Z}$ not a square. (Rational means that one looks for $x, y \in \mathbb{Q}$.) In chapter 20 we will see how that works. Hensel wrote influential books on number theory elaborating his ideas on p-adic numbers. Helmut Hasse was a pupil of Hensel.

approach the verification of the rules of arithmetic is complicated. Anyway, it is a lot of work and it is quite boring. However, for $\mathbb{Z}_p$ it is different: addition and multiplication is done from right to left.

18.16 Example. We take $p = 3$. Suppose we know the first 5 digits of the 3-adic expansion of the 3-adic numbers α and β, say $\alpha = \ldots 10211$ and $\beta = \ldots 11021$. Then we can compute the first 5 digits of the 3-adic expansion of their sum and product:

$$\begin{array}{r} \ldots 10211 \\ \ldots 11021 \\ \hline \ldots 22002 \end{array} \qquad \begin{array}{r} \ldots 10211 \\ \ldots 11021 \\ \hline \ldots 10211 \\ \ldots 1122 \\ \ldots 000 \\ \ldots 11 \\ \ldots 1 \\ \hline \ldots 20201 \end{array}$$

We have introduced $\mathbb{Z}_p$ in an analytic way, namely using limits. We will give another description, an algebraic one, in subsection 18.3.3.

18.3.2 Modular arithmetic in $\mathbb{Z}_p$

In $\mathbb{Z}_p$ one can do arithmetic modulo a power of p. Elements $\alpha \in \mathbb{Z}_p$ can be written as $p^n \mu$ with $\mu \in \mathbb{Z}_p^*$. Arithmetic modulo α then comes down to arithmetic modulo p^n. That is why we do arithmetic modulo powers of p only.

18.17 Definition. Let $\alpha, \beta \in \mathbb{Z}_p$ and $n \in \mathbb{N}^+$. Then we define

$$\alpha \equiv \beta \pmod{p^n} \iff \frac{\alpha - \beta}{p^n} \in \mathbb{Z}_p.$$

We then call α *congruent to* β *modulo* p^n.

Congruence in $\mathbb{Z}_p$ is closely connected to the absolute value:

18.18 Lemma. *Let α and β be elements of $\mathbb{Z}_p$ and $n \in \mathbb{N}^+$. Then:*

$$\alpha \equiv \beta \pmod{p^n} \iff |\alpha - \beta|_p \leq \frac{1}{p^n}.$$

Proof. We have:

$$\alpha \equiv \beta \pmod{p^n} \iff \frac{\alpha-\beta}{p^n} \in \mathbb{Z}_p \iff \left|\frac{\alpha-\beta}{p^n}\right|_p \leq 1 \iff |\alpha-\beta|_p \leq \frac{1}{p^n}. \quad \square$$

It is easily verified that congruence modulo p^n is an equivalence relation and that the equivalence classes form a ring. As with modular arithmetic in $\mathbb{Z}$ addition and multiplication are done on the level of representatives. The ring thus obtained is isomorphic to $\mathbb{Z}/p^n$. We make this more explicit.

For each $\alpha \in \mathbb{Z}_p$ and each $n \in \mathbb{N}^+$ there is an $a_n \in \mathbb{Z}$ with $|\alpha - a_n|_p \leq \frac{1}{p^n}$, that is $\alpha \equiv a_n \pmod{p^n}$. For example $a_n = \sum_{k=0}^{n-1} c_k p^k$, where (c_n) is the p-adic expansion of α. For this a_n we have that $a_n \in \mathbb{N}_{p^n}$. If also an a'_n satisfies, then $a'_n \equiv a_n \pmod{p^n}$. So we have for every n a map $\theta_n \colon \mathbb{Z}_p \to \mathbb{Z}/p^n$ with $\theta_n(\alpha) = \overline{a_n}$, where $a_n \in \mathbb{Z}_p$ such that $|\alpha - a_n|_p \leq \frac{1}{p^n}$.

If (c_n) is the p-adic expansion of $\alpha \in \mathbb{Z}_p$, then $\theta_n(\alpha) = \left(\sum_{k=0}^{n-1} \overline{c_k p^k}\right)$.

18.19 Lemma. *For all $\alpha, \beta \in \mathbb{Z}_p$*

$$\theta_n(\alpha) = \theta_n(\beta) \iff \alpha \equiv \beta \pmod{p^n}.$$

Proof. Choose $a_n, b_n \in \mathbb{Z}$ with $\alpha \equiv a_n \pmod{p^n}$ and $\beta \equiv b_n \pmod{p^n}$. Then

$$\theta_n(\alpha) - \theta_n(\beta) \equiv (\alpha - a_n) - (\beta - b_n) \equiv a_n - b_n \pmod{p^n}. \quad \square$$

18.3.3 Yet another description of $\mathbb{Z}_p$

For every $n \in \mathbb{N}^+$ we have a map $\pi_n \colon \mathbb{Z}/p^{n+1} \to \mathbb{Z}/p^n$, defined by $\pi_n(\overline{a}) = \overline{a}$, where the first $\overline{a}$ is the residue class of $a \in \mathbb{Z}$ modulo p^{n+1} and the second one is the class modulo p^n. Thus we have a sequence of maps

$$\cdots \to \mathbb{Z}/p^{n+1} \xrightarrow{\pi_n} \mathbb{Z}/p^n \to \cdots \to \mathbb{Z}/p^2 \xrightarrow{\pi_1} \mathbb{Z}/p.$$

We will construct a ring $\mathbb{Z}'_p$. The set $\mathbb{Z}'_p$ consists of all sequences $(x_n) = (\ldots, x_{n+1}, x_n, \ldots, x_2, x_1)$, with $x_n \in \mathbb{Z}/p^n$ and $\pi_n(x_{n+1}) = x_n$ for all $n \in \mathbb{N}^+$. In a different notation:

$$\cdots \mapsto x_{n+1} \overset{\pi_n}{\mapsto} x_n \mapsto \cdots \mapsto x_2 \overset{\pi_1}{\mapsto} x_1.$$

Addition and multiplication is done term-wise: $(x_n) + (y_n) = (x_n + y_n)$ and $(x_n) \cdot (y_n) = (x_n y_n)$. Thus we use the addition and multiplication in each of the rings $\mathbb{Z}/p^n$.

We will show that the ring $\mathbb{Z}'_p$ is isomorphic to $\mathbb{Z}_p$. For this we use the maps $\theta_n \colon \mathbb{Z}_p \to \mathbb{Z}/p^n$ from the previous subsection. For every α and every $n \in \mathbb{N}^+$ we have $\theta_n(\alpha) = \overline{a_n}$, where $a_n \in \mathbb{Z}$ with $\alpha \equiv a_n \pmod{p^n}$. Since also $\alpha \equiv a_{n+1} \pmod{p^n}$ we have $\pi_n(\overline{a_{n+1}}) = \overline{a_n}$, that is $\pi_n(\theta_{n+1}(\alpha)) = \theta_n(\alpha)$, for every $n \in \mathbb{N}^+$. Thus we have a map

$$\theta \colon \mathbb{Z}_p \to \mathbb{Z}'_p,\ \alpha \mapsto (\theta_n(\alpha)).$$

18.20 Theorem. *The map $\theta \colon \mathbb{Z}_p \to \mathbb{Z}'_p$ is an isomorphism of rings.*

Proof. Fist we prove the surjectivity of θ. Let $(x_n) \in \mathbb{Z}'_p$. Choose for every $n \in \mathbb{N}^+$ an $a_n \in \mathbb{Z}$ with $\overline{a_n} = x_n$. Then $p^n \mid a_{n+1} - a_n$ and so $(a_{n+1} - a_n)$ is a p-adic null sequence. So (a_n) converges in $\mathbb{Z}_p$, say to α. We prove that $\theta(\alpha) = (x_n)$. Let $n \in \mathbb{N}$. Take an $N > n$ such that $|\alpha - a_N|_p < \frac{1}{p^n}$. Then $|\alpha - a_n|_p \le \max(|\alpha - a_N|_p, |a_N - a_n|_p) \le \frac{1}{p^n}$. From this it follows that $\theta(\alpha) = (\overline{a_n}) = (x_n)$.

Now we prove the injectivity of θ. Let α and β be elements of $\mathbb{Z}_{p^n}$ with $\theta(\alpha) = \theta(\beta)$. Choose for every $n \in \mathbb{N}^+$ an $a_n \in \mathbb{Z}$ and a $b_n \in \mathbb{Z}$ with $\alpha \equiv a_n \pmod{p^n}$ and $\beta \equiv b_n \pmod{p^n}$. Then $\overline{a_n} = \overline{b_n} \in \mathbb{Z}/p^n$ for all $n \in \mathbb{N}^+$. So we could have chosen $a_n = b_n$. Now we have that α and β both are the limit of the sequence (a_n).

The preservation of addition and multiplication by θ is simple: choose numbers a_n for α and b_n for β, then $a_n + b_n$ can be chosen for $\alpha + \beta$ and $a_n b_n$ for $\alpha\beta$. □

18.4 Exponential Functions

18.21 Notations. For $n \in \mathbb{N}$

$$\begin{aligned} p^n\mathbb{Z}_p &= \{\, \alpha \in \mathbb{Z}_p \mid |\alpha|_p \le \tfrac{1}{p^n} \,\} = \{\, \alpha \in \mathbb{Z}_p \mid \alpha \equiv 0 \pmod{p^n} \,\} \\ &= \{\, p^n\alpha \mid \alpha \in \mathbb{Z}_p \,\} = \{\, \alpha \in \mathbb{Z}_p \mid \theta_n(\alpha) = \overline{0} \,\} \end{aligned}$$

and for $n \in \mathbb{N}^+$

$$\begin{aligned} \mathbb{Z}_p^{(n)} &= \{\, \alpha \in \mathbb{Z}_p \mid |\alpha - 1|_p \le \tfrac{1}{p^n} \,\} = \{\, 1 + \alpha \mid \alpha \in p^n\mathbb{Z}_p \,\} \\ &= \{\, \alpha \in \mathbb{Z}_p \mid \alpha \equiv 1 \pmod{p^n} \,\} = \{\, 1 + p^n\alpha \mid \alpha \in \mathbb{Z}_p \,\} \\ &= \{\, \alpha \in \mathbb{Z}_p \mid \theta_n(\alpha) = \overline{1} \,\}. \end{aligned}$$

So if (c_n) is the p-adic expansion of α, then

$$\alpha \in p^n\mathbb{Z}_p \iff c_0 = c_1 = \cdots = c_{n-1} = 0$$

and

$$\alpha \in \mathbb{Z}_p^{(n)} \iff c_0 = 1,\ c_1 = c_2 = \cdots = c_{n-1} = 0.$$

The set $p^n\mathbb{Z}_p$ is closed under addition and under multiplication by elements of $\mathbb{Z}_p$. The set $\mathbb{Z}_p^{(n)}$ is closed under multiplication and inversion. The inverse of $1+\alpha \in \mathbb{Z}_p^{(n)}$ is $1-\alpha+\alpha^2-\alpha^3+\cdots$.

For odd p and $\mu \in \mathbb{Z}_p^{(1)}$ we will define an 'exponential' function $\mathbb{Z}_p \to \mathbb{Z}_p,\ x \mapsto \mu^x$. Also for $p=2$ a function $\mathbb{Z}_2 \to \mathbb{Z}_2,\ x \mapsto \mu^x$ will be defined, however, with the restriction that $\mu \in \mathbb{Z}_2^{(2)}$.

18.22 Lemma. *Let $\mu \in \mathbb{Z}_p$ with $|\mu-1|_p = \frac{1}{p^k}$, where $k \geq 1$ if p odd and $k \geq 2$ if $p=2$. Then $|\mu^p - 1|_p = \frac{1}{p^{k+1}}$.*

Proof. We have $\mu = 1+p^k\alpha$ with $|\alpha|_p = 1$. The binomial formula yields

$$\mu^p = \sum_{l=0}^{p} \binom{p}{l} p^{kl}\alpha^l.$$

Hence

$$\mu^p - 1 - p^{k+1}\alpha = \sum_{l=2}^{p} \binom{p}{l} p^{kl}\alpha^l.$$

For $2 \leq l \leq p-1$ we have $\binom{p}{l}p^{kl}\alpha^l \in p^{k+2}\mathbb{Z}_p$. Also $p^{kp}\alpha \in p^{k+2}\mathbb{Z}_p$ under the given conditions for k. From this it follows that $|\mu^p - 1|_p = \frac{1}{p^{k+1}}$. □

18.23 Lemma. *Let $\mu \in \mathbb{Z}_p^{(1)}$ if p is odd and otherwise $\mu \in \mathbb{Z}_2^{(2)}$. Let (x_n) be a p-adic null sequence in $\mathbb{Z}$. Then $\lim_n \mu^{x_n} = 1$.*

Proof. Let $|\mu-1|_p = \frac{1}{p^k}$ and $n \in \mathbb{N}$. Put $x_n = p^{k_n}y$ with $p \nmid y$. For $y > 0$ it follows from $\mu^y - 1 = (\mu-1)(\mu^{y-1}+\cdots+\mu+1)$ and $\mu \equiv 1 \pmod p$ that $|\mu^y - 1|_p = \frac{1}{p^k}$. This holds for $y<0$ as well: use $\mu^y - 1 = \mu^y(1-\mu^{-y})$. Then by lemma 18.22 we have $|\mu^{x_n} - 1|_p = \frac{1}{p^{k+k_n}}$. Since (x_n) is a p-adic null sequence, it follows that $(\mu^{x_n}-1)$ is a null sequence in $\mathbb{Q}_p$. □

18.24 Proposition. *Let $\mu \in \mathbb{Z}_p^{(1)}$ if p is odd and otherwise $\mu \in \mathbb{Z}_2^{(2)}$. Let $x \in \mathbb{Z}_p$ and let (x_n) be a sequence in $\mathbb{Z}$ converging in $\mathbb{Q}_p$ to x. Then*

(i) *The sequence (μ^{x_n}) converges in $\mathbb{Q}_p$.*
(ii) $|\lim_n \mu^{x_n} - 1|_p = |\mu-1|_p|x|_p$.
(iii) *If also (y_n) is a sequence in $\mathbb{Z}$ converging to x, then the limit of (μ^{y_n}) is equal to the limit of (μ^{x_n}).*

Proof.

(i) Since $(x_{n+1} - x_n)$ is a p-adic null sequence in $\mathbb{Z}$, it follows from lemma 18.23 that $(\mu^{x_{n+1}-x_n} - 1)$ is a p-adic null sequence. From

$$|\mu^{x_{n+1}} - \mu^{x_n}|_p = |\mu^{x_n}|_p |\mu^{x_{n+1}-x_n} - 1|_p = |\mu^{x_{n+1}-x_n} - 1|_p$$

follows that $(\mu^{x_{n+1}} - \mu^{x_n})$ is a null sequence. So (μ^{x_n}) converges in $\mathbb{Q}_p$.

(ii) This follows from lemma 18.22 for $x \neq 0$ and from lemma 18.23 for $x = 0$.

(iii) Since $(x_n - y_n)$ is a p-adic null sequence in $\mathbb{Z}$ and $\mu^{x_n} = \mu^{x_n - y_n}\mu^{y_n}$ the sequences (μ^{x_n}) and (μ^{y_n}) have the same limit. □

This proposition implies that we can define μ^x as follows:

18.25 Definition. Let $\mu \in \mathbb{Z}_p^{(1)}$ if p is odd and otherwise $\mu \in \mathbb{Z}_2^{(2)}$ and let $x \in \mathbb{Z}_p$. Then we define μ^x as the limit of the sequence (μ^{x_n}), where (x_n) is a sequence in $\mathbb{Z}$ converging to x in $\mathbb{Q}_p$.

Thus we have for odd p and every $\mu \in \mathbb{Z}_p^{(1)}$ a map $\mathbb{Z}_p \to \mathbb{Z}_p^{(1)}$, $x \mapsto \mu^x$. And we have for every $\mu \in \mathbb{Z}_2^{(2)}$ a map $\mathbb{Z}_2 \to \mathbb{Z}_2^{(2)}$. These *exponential functions* satisfy the rules one expects:

18.26 Proposition. *For $\mu, \nu \in \mathbb{Z}_p^{(1)}$ and p odd, or $\mu, \nu \in \mathbb{Z}_2^{(2)}$, and $x, y \in \mathbb{Z}_p$:*

(i) $\mu^{x+y} = \mu^x \mu^y$,
(ii) $\mu^x \nu^x = (\mu\nu)^x$.
(iii) $\mu^{xy} = (\mu^x)^y$.

Proof. Choose sequences (x_n) and (y_n) in $\mathbb{Z}$ converging p-adically to x and y respectively.

(i) $\mu^x \mu^y = \lim_n \mu^{x_n} \cdot \lim_n \mu^{y_n} = \lim_n \mu^{x_n + y_n} = \mu^{x+y}$.
(ii) $\mu^x \nu^x = \lim_n \mu^{x_n} \cdot \lim_n \nu^{x_n} = \lim_n (\mu\nu)^{x_n} = (\mu\nu)^x$,
(iii) $(\mu^x)^y = \lim_n (\mu^x)^{y_n}$ and $\mu^{xy} = \lim_n \mu^{x_n y_n}$. From lemma 18.23 follows

$$\frac{(\mu^x)^y}{\mu^{xy}} = \lim_n \frac{(\mu^x)^{y_n}}{\mu^{x_n y_n}} = \lim_n \mu^{(x - x_n)y_n} = 1.$$ □

18.27 Proposition. *For $\mu \in \mathbb{Z}_p^{(1)}$ and p odd the map $\mathbb{Z}_p \to \mathbb{Z}_p^{(1)}$, $x \mapsto \mu^x$ is injective if $\mu \neq 1$. For $\mu \in \mathbb{Z}_2^{(2)}$ the map $\mathbb{Z}_2 \to \mathbb{Z}_2^{(2)}$, $x \mapsto \mu^x$ is injective if $\mu \neq 1$.*

Proof. Suppose that for $x, y \in \mathbb{Z}_p$ we have $\mu^x = \mu^y$. Then $\mu^{x-y} = 1$ and so by proposition 18.26(ii): $|\mu - 1|_p |x - y|_p = |\mu^{x-y} - 1|_p = 0$. If $\mu \neq 1$, then $|\mu - 1|_p \neq 0$ and so $|x - y|_p = 0$, that is $x = y$. □

18.5 The Group $\mathbb{Q}_p^*$

Let $\alpha \in \mathbb{Q}_p^*$. Then $|\alpha|_p = \frac{1}{p^n}$ for an $n \in \mathbb{Z}$ and so $|p^{-n}\alpha|_p = 1$, that is $p^{-n}\alpha \in \mathbb{Z}_p^*$. An element of $\mathbb{Q}_p^*$ can be uniquely written as $p^n\beta$ with $\beta \in \mathbb{Z}_p^*$. Thus we have a group isomorphism $\mathbb{Z} \times \mathbb{Z}_p^* \to \mathbb{Q}_p^*, (n, x) \mapsto p^n x$. The group operation in $\mathbb{Z} \times \mathbb{Z}_p^*$ is addition in the factor $\mathbb{Z}$ and multiplication in $\mathbb{Z}_p^*$. We now focus on $\mathbb{Z}_p^*$.

18.5.1 Roots of unity

See definition 13.44 for the notion of root of unity. We will determine the roots of unity of $\mathbb{Q}_p$. Roots of unity are elements of $\mathbb{Z}_p^*$: if $\zeta^n = 1$, then $|\zeta|_p^n = 1$ and so $|\zeta|_p = 1$. First we show that certain roots of unity can not exist.

18.28 Lemma. *Let $m \in \mathbb{N}$ with $p \mid m$ if p odd and $4 \mid m$ if $p = 2$. Then there is no primitive m-th root of unity in $\mathbb{Q}_p$.*

Proof. Let p odd. It suffices to show that $\mathbb{Q}_p$ has no primitive p-th root of unity.

> Suppose $\zeta \in \mathbb{Q}_p$ is a primitive p-th root of unity. Then ζ is a zero of $x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$. The numbers $\zeta, \zeta^2, \ldots, \zeta^{p-1}$ are the $p-1$ different zeros $\neq 1$ of $x^p - 1$. So
>
> $$x^{p-1} + x^{p-2} + \cdots + x + 1 = (x - \zeta)(x - \zeta^2) \cdots (x - \zeta^{p-1}).$$
>
> For $x = 1$ this yields $p = (1-\zeta)(1-\zeta^2)\cdots(1-\zeta^{p-1})$. Let $c \in \mathbb{Z}$ with $\zeta \equiv c \pmod{p}$. Then by Fermat's Little Theorem: $c \equiv c^p \equiv \zeta^p \equiv 1 \pmod{p}$. So we have $1 - \zeta \equiv 0 \pmod{p}$, that is $\zeta \equiv 1 \pmod{p}$. So also $1 - \zeta^k \equiv 0 \pmod{p}$ for $k \in \mathbb{N}_p$. It follows that $p = (1-\zeta)(1-\zeta^2)\cdots(1-\zeta^{p-1}) \equiv 0 \pmod{p^{p-1}}$. However, $p \not\equiv 0 \pmod{p^{p-1}}$. Contradiction.

We prove that $\mathbb{Q}_2$ has no primitive 4-th root of unity.

> Suppose $\zeta \in \mathbb{Q}_2$ is a primitive 4-th root of unity. Then ζ and $-\zeta$ are different zeros of x^2+1. So $x^2+1 = (x-\zeta)(x+\zeta)$. For $x = 1$ this gives $2 = (1-\zeta)(1+\zeta)$. Since $\zeta \in \mathbb{Z}_2^*$, we have $\zeta \equiv 1 \pmod{2}$. So $2 \equiv 0 \pmod{4}$. Contradiction. □

We will show that $\mathbb{Q}_p$ has a primitive $(p-1)$-st root of unity.

18.29 Lemma. *Let $\alpha \in \mathbb{Z}_p^*$. Then the sequence (α^{p^n}) converges in $\mathbb{Q}_p$ to a $(p-1)$-st root of unity.*

Proof. For convergence of (α^{p^n}) it suffices to show that the difference sequence $(\alpha^{p^{n+1}} - \alpha^{p^n})$ is a null sequence. We have

$$|\alpha^{p^{n+1}} - \alpha^{p^n}|_p = |\alpha^{p^n}|_p |(\alpha^{p-1})^{p^n} - 1|_p = |(\alpha^{p-1})^{p^n} - 1|_p.$$

Let $\alpha \equiv a \pmod p$ with $a \in \mathbb{Z}$. Then by Fermat's Little Theorem $\alpha^{p-1} \equiv a^{p-1} \equiv 1 \pmod p$, that is $\alpha^{p-1} \in \mathbb{Z}_p^{(1)}$. For p odd lemma 18.23 implies that $((\alpha^{p-1})^{p^n})$ converges to 1. So $(\alpha^{p^{n+1}} - \alpha^{p^n})$ is a null sequence. For $p = 2$ we have $(\alpha^{p-1})^{p^n} = \alpha^{2^n} = (\alpha^2)^{2^{n-1}}$ with $\alpha^2 \in \mathbb{Z}_2^{(2)}$. Also in this case lemma 18.23 can be applied.

Let ζ be the limit of the sequence (α^{p^n}). From $\alpha^{p^{n+1}} = (\alpha^{p^n})^p$ follows that $\zeta = \zeta^p$. Because $\alpha^{p^n} \in \mathbb{Z}_p^*$ for all n, also ζ is an element of $\mathbb{Z}_p^*$. In particular ζ is not 0. So $\zeta^{p-1} = 1$. □

18.30 Definition. We define the map $\omega\colon \mathbb{Z}_p^* \to \mathbb{Z}_p^*$ by $\omega(\alpha) = \lim_n \alpha^{p^n}$.

18.31 Lemma. *For all $\alpha, \beta \in \mathbb{Z}_p^*$:*

(i) $\omega(\alpha\beta) = \omega(\alpha)\omega(\beta)$.
(ii) $\omega(\alpha) \equiv \alpha \pmod p$.
(iii) $\omega(\alpha) = \omega(\beta) \iff \alpha \equiv \beta \pmod p$.

Proof.

(i) This follows from $(\alpha\beta)^{p^n} = \alpha^{p^n}\beta^{p^n}$.
(ii) From $|\alpha^p - \alpha|_p < 1$ follows that $|\alpha^{p^n} - \alpha|_p < 1$ for all $n \in \mathbb{N}$. So $|\omega(\alpha) - \alpha|_p < 1$, that is $\omega(\alpha) \equiv \alpha \pmod p$.
(iii) "⇒" follows from (ii). Suppose $\alpha \equiv \beta \pmod p$. Then $\frac{\alpha}{\beta} \in \mathbb{Z}_p^{(1)}$. For p odd $((\frac{\alpha}{\beta})^{p^n})$ converges by lemma 18.23 to 1, because (p^n) is a p-adic null sequence $\mathbb{Z}$. So $\omega(\frac{\alpha}{\beta}) = 1$ and using part (i): $\omega(\alpha) = \omega(\beta)$. For $p = 2$ we have $\omega(\alpha) = \omega(\beta) = 1$. □

By lemma 18.31 ω induces an injective map $\mathbb{F}_p^* \to \mathbb{Z}_p^*$, $\overline{a} \mapsto \omega(a)$. If g is a primitive root modulo p, then is $\omega(g)$ a primitive $(p-1)$-st root of unity of $\mathbb{Q}_p^*$.

18.32 Notation. The injective map $\mathbb{F}_p^* \to \mathbb{Z}_p^*$ induced by ω we also denote by ω.

For all $\overline{a}, \overline{b} \in \mathbb{F}_p^*$ we have $\omega(\overline{ab}) = \omega(\overline{a})\omega(\overline{b})$. Now we have a group isomorphism $\mathbb{F}_p^* \times \mathbb{Z}_p^{(1)} \to \mathbb{Z}_p^*$, $(\overline{k}, x) \mapsto \omega(k)x$. In the next subsection we will focus on $\mathbb{Z}_p^{(1)}$.

18.5.2 The group $\mathbb{Z}_p^{(1)}$

First we consider odd p. By proposition 18.27 the map $\mathbb{Z}_p \to \mathbb{Z}_p^{(1)}$, $x \mapsto \mu^x$ is injective for $\mu \in \mathbb{Z}_p^{(1)}$. Now we will determine the image of this exponential function.

18.33 Theorem. *Let p be odd and $\mu \in \mathbb{Q}_p$ with $|\mu - 1|_p = \frac{1}{p}$. Then the exponential map*

$$\mathbb{Z}_p \to \mathbb{Z}_p^{(1)},\ x \mapsto \mu^x$$

is bijective.

Proof. The injectivity of the map has already been shown (proposition 18.27). Let $y \in \mathbb{Z}_p^{(1)}$. To prove that there exists an $x \in \mathbb{Z}_p$ such that $\mu^x = y$.

Take sequences (a_n) and (y_n) in $\mathbb{Z}$ with $\mu \equiv a_n \pmod{p^n}$ and $y \equiv y_n \pmod{p^n}$ for all $n \in \mathbb{N}$.

From proposition 18.24(ii) follows that for $n \in \mathbb{N}^+$ we have

$$|\mu^{p^n} - 1|_p = |\mu - 1|_p |p^n|_p = \frac{1}{p^{n+1}}.$$

So for $n \in \mathbb{N}^+$ we have $a_{n+1}^{p^n} \equiv \mu^{p^n} \equiv 1 \pmod{p^{n+1}}$ and $a_{n+1}^{p^{n-1}} \equiv \mu^{p^{n-1}} \not\equiv 1 \pmod{p^{n+1}}$. It follows that $\mathrm{o}_{p^{n+1}}(a_{n+1}) = p^n$. Hence there are p^n different powers of $\overline{a_{n+1}}$ in $(\mathbb{Z}/p^{n+1})^*$ and these must be all the elements $\overline{a} \in (\mathbb{Z}/p^{n+1})^*$ with $\overline{a} = \overline{1}$ in $\mathbb{Z}/p$.

Since $y \in \mathbb{Z}_p^{(1)}$, we have $\overline{y_{n+1}} = \overline{1}$ in $\mathbb{Z}/p$. So for every $n \in \mathbb{N}^+$ there is an $x_n \in \mathbb{Z}$ with $\overline{a_{n+1}}^{x_n} = \overline{y_{n+1}}$ in $\mathbb{Z}/p^{n+1}$. From $\overline{a_{n+2}}^{x_{n+1}} = \overline{y_{n+2}}$ in $\mathbb{Z}/p^{n+2}$ follows that also $\overline{a_{n+1}}^{x_{n+1}} = \overline{y_{n+1}}$ in $\mathbb{Z}/p^{n+1}$. So since $\mathrm{o}_{p^{n+1}}(a_{n+1}) = p^n$ we have $x_{n+1} \equiv x_n \pmod{p^n}$.

Now take $x = \lim_n(x_n)$. From $\overline{a_{n+1}}^{x_n} = \overline{y_{n+1}}$ in $\mathbb{Z}/p^{n+1}$ follows $|a_{n+1}^{x_n} - y_{n+1}|_p \leq \frac{1}{p^{n+1}}$. Since $|\mu - a_{n+1}|_p \leq \frac{1}{p^{n+1}}$ it follows that $|\mu^{x_n} - y_{n+1}|_p \leq \frac{1}{p^{n+1}}$. The sequences (μ^{x_n}) and (y_n) differ by a null sequence. So $\mu^x = y$. □

For $p = 2$ we have:

18.34 Theorem. *Let $\mu \in \mathbb{Q}_2$ with $|\mu - 1|_2 = \frac{1}{4}$. Then the exponential map*

$$\mathbb{Z}_2 \to \mathbb{Z}_2^{(2)},\ x \mapsto \mu^x$$

is bijective.

Proof. The proof is analogous to the one of theorem 18.33. Since $|\mu - 1|_2 = \frac{1}{4}$ we now have $|\mu^{2^n} - 1|_2 = \frac{1}{2^{n+2}}$. Furthermore, we now look at elements of $\overline{a} \in \mathbb{Z}/2^{n+2}$ with $\overline{a} = \overline{1}$ in $\mathbb{Z}/4$. There are 2^n of these. □

18.35 Corollary. *Let $n \in \mathbb{N}^+$ and assume that $n \geq 2$ if $p = 2$. Let $\mu \in \mathbb{Z}_p^{(n)}$. Then $\mathbb{Z}_p^{(n)}$ is the image of the exponential map*

$$\mathbb{Z}_p \to \mathbb{Z}_p^{(1)},\ x \mapsto \mu^x.$$

Proof. We have already seen that the image is contained in $\mathbb{Z}_p^{(n)}$. Let $y \in \mathbb{Z}_p^{(n)}$. To prove that there is an $x \in \mathbb{Z}_p$ with $\mu^x = y$.

If p is odd, then choose a μ_1 with $|\mu_1 - 1|_p = \frac{1}{p}$. By theorem 18.33 there are $x_1, x_2 \in \mathbb{Z}_p$ with $\mu_1^{x_1} = \mu$ and $\mu_1^{x_2} = y$. Then $\mu^{\frac{x_2}{x_1}} = (\mu_1^{x_1})^{\frac{x_2}{x_1}} = \mu^{x_2} = y$.

For $p = 2$ choose a $\mu \in \mathbb{Z}_2^{(2)}$. Now use theorem 18.34. □

18.5.3 The structure of $\mathbb{Q}_p^*$

The importance of the theorems 18.33 and 18.34 is that the seemingly complicated structure of the multiplicative group $\mathbb{Z}_p^{(1)}$ is translated into the simple structure of the additive group $\mathbb{Z}_p$. As for $\mathbb{R}$ the multiplicative structure of $\mathbb{Q}_p$ is isomorphic to an additive one.

18.36 Theorem. *Let p be an odd prime number, $g \in \mathbb{Z}$ a primitive root modulo p and $\alpha \in \mathbb{Q}_p^*$. Then there are unique $n \in \mathbb{N}$, $\overline{k} \in \mathbb{Z}/(p-1)$ and $x \in \mathbb{Z}_p$ such that*

$$\alpha = p^n \omega(g)^k (1+p)^x.$$

Proof. If $|\alpha|_p = \frac{1}{p^n}$, then $\alpha_0 = p^{-n}\alpha \in \mathbb{Z}_p^*$. There is a unique $\overline{k} \in \mathbb{Z}/(p-1)$ with $\alpha_0 \equiv g^k \pmod{p}$. Then $\alpha_1 = \alpha_0 \omega(g)^{-k} \in \mathbb{Z}_p^{(1)}$. Since $|p|_p = \frac{1}{p}$, there finally is a unique $x \in \mathbb{Z}_p$ with $(1+p)^x = \alpha_1$. □

Multiplication now becomes addition of the exponents: if $\alpha = p^n \omega(g)^k (1+p)^x$ and $\beta = p^m \omega(g)^l (1+p)^y$, then $\alpha\beta = p^{n+m} \omega(g)^{k+l} (1+p)^{x+y}$. So multiplication in $\mathbb{Q}_p^*$ can for odd p be translated into component-wise addition in

$$\mathbb{Z} \times \mathbb{Z}/(p-1) \times \mathbb{Z}_p.$$

So we have a group isomorphism

$$\mathbb{Z} \times \mathbb{Z}/(p-1) \times \mathbb{Z}_p \to \mathbb{Q}_p^*,\ (n, \overline{k}, x) \mapsto p^n \omega(g)^k (1+p)^x.$$

18.37 Theorem. *Let $\alpha \in \mathbb{Q}_2^*$. Then there are unique $n \in \mathbb{Z}$, $\overline{k} \in \mathbb{Z}/2$ and $x \in \mathbb{Z}_2$ such that*

$$\alpha = 2^n (-1)^k 5^x.$$

Proof. Again we have $\alpha_0 = 2^{-n}\alpha \in \mathbb{Z}_2^*$. There is a unique $\overline{k} \in \mathbb{Z}/2$ with $(-1)^k \alpha_0 \equiv 1 \pmod{4}$. Then $\alpha_1 = (-1)^k \alpha_0 \in \mathbb{Z}_2^{(2)}$. Since $|5-1|_2 = \frac{1}{4}$ there is a unique $x \in \mathbb{Z}_2$ with $5^x = \alpha_1$. □

Here too multiplication becomes addition of exponents: if $\alpha = 2^n (-1)^k 5^x$ and $\beta = 2^m (-1)^l 5^y$, then $\alpha\beta = 2^{n+m} (-1)^{k+l} 5^{x+y}$. So multiplication can be translated into component-wise addition

$$\mathbb{Z} \times \mathbb{Z}/2 \times \mathbb{Z}_2.$$

We have a group isomorphism

$$\mathbb{Z} \times \mathbb{Z}/2 \times \mathbb{Z}_2 \to \mathbb{Z}_2^*,\ (n, \overline{k}, x) \mapsto 2^n (-1)^k 5^x.$$

18.5.4 Powers

Let $m \in \mathbb{N}^+$ and p an odd prime. What are the m-th powers in $\mathbb{Q}_p^*$? For this we use the description of $\mathbb{Q}_p^*$ derived in the previous subsection. The m-th power of $\alpha = p^n\omega(g)^k(1+p)^x$ is

$$\alpha^m = p^{nm}\omega(g)^{km}(1+p)^{xm}.$$

So an element $\beta = p^N\omega(g)^K(1+p)^X$ with $N \in \mathbb{Z}$, $K \in \mathbb{Z}$ and $X \in \mathbb{Z}_p$ is an m-th power if and only if

a) N is a multiple of m,
b) $K \equiv km \pmod{p-1}$ for some $k \in \mathbb{Z}$, that is, there are $k, l \in \mathbb{Z}$ with $K = km + l(p-1)$, which comes down to K being a multiple of $\gcd(m, p-1)$,
c) $\frac{X}{m} \in \mathbb{Z}_p$, that is, $|X|_p \leq |m|_p$. Put $t = v_p(m)$. So the condition is $|X|_p \leq \frac{1}{p^t}$, that is $X \in p^t\mathbb{Z}_p$.

If the conditions are satisfied, then the number of $\alpha \in \mathbb{Q}_p^*$ with $\alpha^m = \beta$ is equal to $\gcd(m, p-1)$.

Now the case $p = 2$. Let $m \in \mathbb{N}^+$. The m-th power of $\alpha = 2^n(-1)^k5^x$ is

$$\alpha^m = 2^{nm}(-1)^{km}5^{xm}.$$

So an element $\beta = 2^N(-1)^K5^X$ with $N \in \mathbb{Z}$, $K \in \mathbb{Z}$ and $X \in \mathbb{Z}_2$ is an m-th power if and only if

a) N is a multiple of m,
b) $K \equiv km \pmod{2}$ for a $k \in \mathbb{Z}$, that is, K is even if m is even,
c) $\frac{X}{m} \in \mathbb{Z}_2$, that is $|X|_p \leq |m|_2$. Put $t = v_2(m)$. So the condition is $|X|_2 \leq \frac{1}{2^t}$, that is $X \in 2^t\mathbb{Z}_2$.

If the conditions are satisfied, then the number of α with $\alpha^m = \beta$ is $\gcd(m, 2)$.

18.5.5 The multiplicative group modulo squares

We use the notations of the previous subsection. Let's look at squares, that is $m = 2$. They form the set

$$\mathbb{Q}_p^{*2} = \{\, \alpha^2 \mid \alpha \in \mathbb{Q}_p^* \,\}.$$

For p odd we get: $\beta \in \mathbb{Q}_p^*$ is a square if and only if

a) N is even,
b) K is even,
c) $v_p(2) = 0$, so any X satisfies.

So there are four types of elements of $\mathbb{Q}_p^*$:

$$\alpha^2, \quad p\alpha^2, \quad \omega(a)\alpha^2, \quad p\omega(a)\alpha^2,$$

where $\alpha \in \mathbb{Q}_p^*$ and $a \in \mathbb{Z}$ with $\left(\frac{a}{p}\right) = -1$. Since $\frac{\omega(a)}{a} \in \mathbb{Z}_p^{(1)}$ and therefore a square, the set $\mathbb{Q}_p^*$ is the disjoint union of

$$\mathbb{Q}_p^{*2}, \quad p\mathbb{Q}_p^{*2}, \quad b\mathbb{Q}_p^{*2} \quad \text{and} \quad pb\mathbb{Q}_p^{*2}.$$

where $b \in \mathbb{Z}$, a nonsquare modulo p. It are the equivalence classes of the equivalence relation $\sim$ in $\mathbb{Q}_p^*$ given by $\alpha \sim \beta \iff \frac{\alpha}{\beta} \in \mathbb{Q}_p^{*2}$.

18.38 Example. The integer 3 is a nonsquare modulo 7, so the four classes in $\mathbb{Q}_7^*$ are

$$\mathbb{Q}_7^{*2}, \quad 7\mathbb{Q}_7^{*2}, \quad 3\mathbb{Q}_7^{*2} \quad \text{and} \quad 21\mathbb{Q}_7^{*2}.$$

The group $\mathbb{Q}^*$ has infinitely many classes modulo squares, but here we have only four. Note that 2 is not a square in $\mathbb{Q}$, but is a square in $\mathbb{Q}_7^*$: $\frac{2}{\omega(2)} \in \mathbb{Z}_7^{(1)}$, by theorem 18.33 all elements of $\mathbb{Z}_7^{(1)}$ are squares and $\omega(2)$ is a square ($\omega(2) = \omega(3)^2$).

For $p = 2$ the number $\beta \in \mathbb{Q}_2^*$ is a square if and only if

a) N is even,
b) K is even,
c) $X \in 2\mathbb{Z}_2$ (since $v_2(2) = 1$).

So there are eight types of elements in $\mathbb{Q}_2^*$:

$$2^i(-1)^j 5^k \alpha^2,$$

where $\alpha \in \mathbb{Q}_2^*$ and $i, j, k \in \{0, 1\}$. In this case there are eight classes modulo squares:

$$\mathbb{Q}_2^{*2}, \quad 2\mathbb{Q}_2^{*2}, \quad -\mathbb{Q}_2^{*2}, \quad 5\mathbb{Q}_2^{*2}, \quad -2\mathbb{Q}_2^{*2}, \quad 10\mathbb{Q}_2^{*2}, \quad -5\mathbb{Q}_2^{*2} \quad \text{and} \quad -10\mathbb{Q}_2^{*2}.$$

18.39 Example. How to determine the class of a given element of $\mathbb{Q}^*$? Let's take the number 7. By theorem 18.34 all elements in the group $\mathbb{Z}_2^{(2)}$ are squares. Because $-7 \in \mathbb{Z}_2^{(2)}$, we have $7 \in -\mathbb{Q}_2^{*2}$.

Exercises

1. The first 8 digits of the 3-adic expansion of the 3-adic number α are known: $\alpha = \ldots 12002112$. What are the first 8 digits of the 3-adic expansion of $1 - \alpha$?

2. The first 3 digits of the 7-adic expansion of $\alpha \in \mathbb{Z}_7$ are $2, 2, 1$, so: $\alpha = \ldots 122$. The 7-adic number $\beta = \lim_n \alpha^{7^n}$ is a 6th root of unity. Is it a primitive 6th root of unity?

3. (i) For which prime numbers p is -1 a square in $\mathbb{Q}_p$?
   (ii) Compute the first three digits of the 5-adic expansion of $\sqrt{-1} \in \mathbb{Q}_5$.

4. Let $r \in \mathbb{Q}^*$. Prove that r is a square in $\mathbb{Q}$ if and only if r is a square in all completions of $\mathbb{Q}$.

5. Let a, b and c be rational numbers with $a \neq 0$. Prove that the quadratic equation $ax^2 + bx + c = 0$ is solvable in $\mathbb{Q}$ if and only if it is so in all completions of $\mathbb{Q}$.

6. Let p be an odd prime number. The numbers $\alpha, \beta \in \mathbb{Z}_p^*$ are not squares in $\mathbb{Q}_p$. Show that $\alpha\beta$ is a square in $\mathbb{Q}_p$.

7. Let p and q be two different prime numbers. Show that there is an $a \in \mathbb{Z}$ such that $\lim_n^{(p)} a^{q^n} = 0$ and $\lim_n^{(q)} a^{q^n} = 1$.

8. From theorem 18.33 it follows that there is a unique $x \in \mathbb{Z}_3$ such that $4^x = 7$ in $\mathbb{Q}_3$. Determine the first three digits of the 3-adic expansion of x.

9. Determine the first three digits of the 2-adic expansion of $x \in \mathbb{Z}_2$ which satisfies $5^x = 9$ in $\mathbb{Q}_2$.

10. Is there a sequence in $\mathbb{Q}$, being a null sequence with respect to the ordinary metric while converging to 1 with respect to the 3-adic metric?

Part V

Extensions

This final part is about three topics, all of them related to square roots of elements in a field:

- The complex numbers are constructed in chapter 19. They are at the end of the contruction of the number system: they contain the field $\mathbb{R}$ and it is shown that any polynomial equation of positive degree has a solution.
- For $a, b \in \mathbb{Q}^*$ and a not a square the equation

  $$x^2 - ay^2 = b.$$

  To find solutions $(x, y) \in \mathbb{Q}^2$. This is completely solved in chapter 20. It uses the results described in subsection 18.5.5 on classes modulo squares for the field of p-adic numbers. This topic is the most advanced one in this book.
- For $d \in \mathbb{N}^+$ and d not a square the Diophantine equation

  $$x^2 - dy^2 = \pm 1.$$

  The equation is known as Pell's equation. It is shown in chapter 21 that for any d the equation has a solution, in fact infinitely many of them. An algorithm for the solution is given.

19 The Complex Numbers

Our starting point was $\mathbb{N}$, the natural numbers. By consecutive extensions of the number system we have achieved that more and more equations have solutions. The equation $x+7=3$ is not solvable in $\mathbb{N}$, but it is in $\mathbb{Z}$, the equation $3x+5=0$ is not solvable in $\mathbb{Z}$, but it is in $\mathbb{Q}$, the equation $x^2-2=0$ is not solvable in $\mathbb{Q}$, but it is in $\mathbb{R}$. In extending from $\mathbb{Q}$ to $\mathbb{R}$ many new numbers emerged, not only solutions of equations, but also transcendental numbers, such as π and e. Still there are equations without a solution in $\mathbb{R}$, for example the equations $x^2+a=0$ with $a>0$.

In section 19.1 we will see that for solving cubic equations square roots of negative numbers are used, if even the solutions themselves are real. This method for the solution of cubic equations was found in Italy more than five centuries ago, so in times one still was struggling with negative numbers. In section 19.2 $\mathbb{R}$ will be extended with the square root of -1, thus obtaining $\mathbb{C}$, the field of complex numbers. As we will see, many more equations will have solutions, in fact all equations have; this is the so-called Fundamental Theorem of Algebra. Moreover, $\mathbb{C}$ is complete: Cauchy sequences do converge. Thus the field $\mathbb{C}$ is a natural end point in a succession of extensions of number systems.

19.1 Cubic Equations

Around 1500 in Italy a method was found for the solution of cubic equations. It was found by Del Ferro, later independently by Fontana (better known as Tartaglia, the stammerer) and published by Cardano. When applied to a general cubic equation, a formula is obtained for its solution. This formula is known as Cardan's Formula. We describe the method by a worked example.

In general an equation $x^3+ax^2+bx+c=0$ can be transformed into one of type $y^3+py+q=0$ by substituting $x=y-\frac{a}{3}$, thus obtaining a cubic equation with no quadratic term. We will solve the equation $x^3-7x-6=0$. In this particular case it is easy to find a solution just by trying. We will not do so since the idea is to demonstrate the method in general.

The crucial step is to write x as a sum of two terms

$$x=u+v$$

Nicolo Fontana (Brescia 1500 – Venice 1557)

He is better known under the nickname Tartaglia. Tartaglia found a method for solving cubic equations. He did so challenged by, the not so brilliant, pupil **Fior** of **Scipione del Ferro** (Bologna 1465 – Bologna 1526) who, as became clear later, had found this method before.

and to substitute this in the equation:

$$(u+v)^3 - 7(u+v) - 6 = 0.$$

Thus we get

$$u^3 + 3uv(u+v) + v^3 - 7(u+v) - 6 = 0.$$

We choose u and v such that $3uv = 7$. Then the two terms containing the factor $(u+v)$ cancel:

$$u^3 + v^3 - 6 = 0.$$

Using $3uv = 7$ gives

$$u^6 - 6u^3 + \frac{7^3}{3^3} = 0,$$

that is

$$3^3u^6 - 6 \cdot 3^3u^3 + 7^3 = 0$$

and this is a quadratic equation in u^3. A solution is

$$u^3 = \frac{6 \cdot 3^3 + 3\sqrt{6^2 \cdot 3^4 - 4 \cdot 7^3 \cdot 3}}{2 \cdot 3^3} = \frac{3^4 + 30\sqrt{-3}}{3^3} = \frac{(-3 + 2\sqrt{-3})^3}{3^3}.$$

So we can take $u = \frac{-3+2\sqrt{-3}}{3}$. We ignored that we do not really have a thing like the square root of -3. We just calculate with $\sqrt{-3}$ as a number having -3 as its square. We now have u and since $3uv = 7$ we also find $v = \frac{-3-2\sqrt{-3}}{3}$. So

$$x = u + v = \frac{-3 + 2\sqrt{-3}}{3} + \frac{-3 - 2\sqrt{-3}}{3} = -1 - 1 = -2$$

and we have found a solution. Though the solution is real, during the computation we have used other 'nonexisting' numbers. Whether these numbers do exist or not, it is a way to find a solution.

Girolimo Cardano (Pavia 1501 – Rome 1576)

Cardano (or Cardan) published in his book *Ars Magna* which was totally devoted to algebra, Tartaglia's method for solving cubic equations. Cardano had promised not to do so, but when he heard of Del Ferro's earlier work he decided to publish the method. Tartaglia was publicly challenged by Cardano's secretary Ferrari and he lost. That was no wonder: **Ludovico Ferrari** (Bologna 1522 – Bologna 1565) appeared to be very gifted and had, on request of Cardano, found a method for solving equations of degree four. The life of Cardano was remarkable. He was well-known as a medical doctor and as such was invited to come to Scotland for treating the archbishop. Later he had a conflict with the pope because of blasphemy, but nevertheless later the pope awarded him a pension.

This method of solving cubic equations shows that it makes sense to extend $\mathbb{R}$ further. The rules of arithmetic have to remain valid in this extension. Thus the use of nonexisting numbers will disappear.

> We have found only one solution. In fact there will be three cubic roots of u and when using each of them one finds three solutions. We will describe this in section 19.4. Application of Tartaglia's method to $x^3+px+q=0$ leads to *Cardan's formula*:
>
> $$x = \sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{p}{3}\right)^3+\left(\frac{q}{2}\right)^2}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{p}{3}\right)^3+\left(\frac{q}{2}\right)^2}}.$$

19.2 Construction of the Complex Numbers

We want to extend $\mathbb{R}$ in such a way that also -1 is a square, say $-1=i^2$, and moreover the rules of arithmetic still hold. Possibly we want too much. A construction of such an extension will show its existence. To conceive such a construction it is as always instructive first to look at consequences of its existence. Since addition and multiplication will exist in the extended system it contains numbers $a+bi$ with $a,b\in\mathbb{R}$. The rules of arithmetic imply that addition and multiplication will not give anything new:

$$\begin{aligned}(a+bi)+(c+di) &= a+c+bi+di=(a+c)+(b+d)i\\ (a+bi)\cdot(c+di) &= ac+adi+bci+bdi^2=(ac-bd)+(ad+bc)i.\end{aligned}$$

The number $a+bi$ is completely determined by $(a,b)\in\mathbb{R}^2$. Can numbers $a+bi$ and $c+di$ be equal while $(a,b)\neq(c,d)$? If $a+bi=c+di$, then $(a-c)+(b-d)i=0$.

If $b - d \neq 0$, then it would follow that $i \in \mathbb{R}$. Contradiction. So $b = d$ and as a consequence $a = c$ as well. The numbers we want correspond to elements of $\mathbb{R}^2$ and this observation will be the basis of the construction. Note that inside the system division is also possible: if $a + bi \neq 0$, and so $(a, b) \neq (0, 0)$, then $(a + bi)(\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i) = 1$.

19.2.1 The construction

We take the set $\mathbb{R}^2$ of ordered pairs of real numbers. On this set we define an addition and a multiplication:

$$\begin{aligned}(a, b) + (c, d) &= (a + c, b + d)\\ (a, b) \cdot (c, d) &= (ac - bd, ad + bc).\end{aligned}$$

19.1 Definition. We denote the set $\mathbb{R}^2$ together with the above defined addition and multiplication by $\mathbb{C}$. The elements of $\mathbb{C}$ are called *complex* numbers.

As a set $\mathbb{R}^2$ and $\mathbb{C}$ coincide. The notation $\mathbb{C}$ stands for the set together with this addition and multiplication.

19.2 Theorem. *$\mathbb{C}$ is a field.*

Proof. It is a matter of straightforward verification that the rules of arithmetic hold for addition and multiplication. The null element is $(0, 0)$, the unit element is $(1, 0)$. For example the distributive law:

$$\begin{aligned}(a, b) \cdot ((c, d) + (e, f)) &= (a, b) \cdot (c + e, d + f)\\ &= (a(c + e) - b(d + f), a(d + f) + b(c + e))\end{aligned}$$

and

$$\begin{aligned}(a, b) \cdot (c, d) + (a, b) \cdot (e, f) &= (ac - bd, ad + bc) + (ae - bf, af + be)\\ &= (ac - bd + ae - bf, ad + bc + af + be).\end{aligned}$$

The existence of inverses is, after the preparations made, not difficult either: if $(a, b) \neq (0, 0)$, then

$$(a, b) \cdot \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right) = (1, 0). \qquad \square$$

19.2.2 $\mathbb{C}$ as extension of $\mathbb{R}$

Arithmetic with the elements $(a, 0)$ corresponds to the arithmetic with the reals. More precisely:

19.3 Proposition. *The map $\mathbb{R} \to \mathbb{C}$, $a \mapsto (a, 0)$ is injective and respects the addition and multiplication: $a + b \mapsto (a, 0) + (b, 0)$ and $ab \mapsto (a, 0) \cdot (b, 0)$.* □

From now on we identify $(a, 0)$ with a and we write i for $(0, 1)$. Then we have

$$(a, b) = (a, 0) + (0, b) = (a, 0) + (b, 0)(0, 1) = a + bi.$$

The square of i is -1, as intended:

$$i^2 = (0, 1)^2 = (-1, 0) = -1.$$

For arithmetic with a variable varying over real numbers it is customary to denote this variable as x. For arithmetic with complex numbers usually a z is used and it is standard to write $z = x + yi$. Then the complex variable z corresponds to two real variables x and y.

19.4 Definition. For $z = x + yi$ a complex number, the x is called the *real part* of z and y the *imaginary part* of z. Notation: $\Re(z) = x$ and $\Im(z) = y$. The number $x - yi$ is called the *(complex) conjugate* of z. Notation: $x - yi = \overline{z}$. The transformation $z \mapsto \overline{z}$ of $\mathbb{C}$ is called *complex conjugation*.

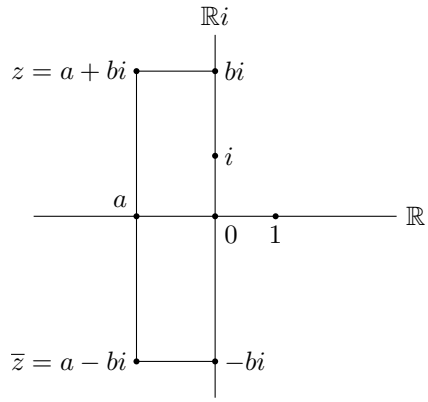


Figure 19.1: The complex plane

19.5 Proposition. *Complex conjugation is an addition and multiplication preserving bijection.*

Proof. This is easy:

$$\begin{aligned}\overline{(a+bi)+(c+di)} &= \overline{(a+c)+(b+d)i} = (a+c)-(b+d)i \\ &= (a-bi)+(c-di) = \overline{a+bi}+\overline{c+di} \\ \overline{(a+bi)\cdot(c+di)} &= \overline{(ac-bd)+(ad+bc)i} = (ac-bd)-(ad+bc)i \\ &= (a-bi)\cdot(c-di) = \overline{a+bi}\cdot\overline{c+di}.\end{aligned}$$

It is a bijection: this is a consequence of the fact that conjugating twice yields the identical transformation. □

So complex conjugation is an isomorphism from $\mathbb{C}$ to itself. A transformation which is an isomorphism is usually called an *automorphism*. Complex conjugation is an automorphism of $\mathbb{C}$. Precisely the real numbers are fixed under this automorphism.

19.2.3 The completeness of $\mathbb{C}$

The field $\mathbb{R}$ is complete: on $\mathbb{R}$ we have an absolute value and Cauchy sequences w.r.t. this absolute value converge. The absolute value on $\mathbb{R}$ is an extension of the ordinary absolute value on $\mathbb{Q}$. It can be extended further to an absolute value on $\mathbb{C}$.

19.6 Definition. The *absolute value* $|z|$ (or *modulus*) of $z = x + yi \in \mathbb{C}$ is

$$|z| = \sqrt{x^2+y^2} = \sqrt{z\cdot\overline{z}}.$$

19.7 Proposition. *The absolute value on $\mathbb{C}$ satisfies the requirements for an absolute value on a field.*

Proof. The identity $|z_1 z_2| = |z_1|.|z_2|$ follows directly from the definition:

$$|z_1 z_2|^2 = z_1 z_2 \cdot \overline{z_1 z_2} = z_1 \cdot \overline{z_1} \cdot z_2 \overline{z_2} = |z_1|^2 \cdot |z_2|^2.$$

The absolute value on $\mathbb{C}$ is the same as the standard metric on $\mathbb{R}^2$, so the remaining part of the proposition follows from proposition 17.39 □

The absolute value on $\mathbb{C}$ comes with the notions of 'converging sequence' and 'Cauchy sequence' in $\mathbb{C}$. We also have the usual rules for limits.

19.8 Theorem. *The field $\mathbb{C}$ is complete.*

Proof. Let (z_n) be a Cauchy sequence in $\mathbb{C}$. We write $z_n = x_n + y_n i$ with x_n and y_n real. From $|x_n - x_m| \leq |z_n - z_m|$ and $|y_n - y_m| \leq |z_n - z_m|$ follows that (x_n) and (y_n) are Cauchy sequences in $\mathbb{R}$. These converge in $\mathbb{R}$, since $\mathbb{R}$ is complete. From the rules for limits follows that (z_n) converges as well. □

So, if (z_n) is a Cauchy sequence in $\mathbb{C}$, then we have

$$\lim_n z_n = \lim_n x_n + \lim_n y_n \cdot i.$$

19.3 The Group $\mathbb{C}^*$

We have seen that for the completions $\mathbb{R}$ and $\mathbb{Q}_p$ the multiplicative groups are closely connected to the additive groups via exponential functions. For $\mathbb{C}$ we have something similar.

19.3.1 The exponential function

The function $\exp\colon \mathbb{R} \to \mathbb{R}$, defined in section 17.6, can be extended to $\mathbb{C}$ by defining it in the same way as was done for $\mathbb{R}$.

Let $z \in \mathbb{C}$. We start with

$$e_n(z) = \sum_{k=0}^{n} \frac{z^k}{k!} \qquad \text{for all } n \in \mathbb{N}$$

and we will show that the sequence $(e_n(z))$ converges.

19.9 Lemma. *The sequence $(e_n(z))$ converges for all $z \in \mathbb{C}$.*

Proof. For $n \geq m \in \mathbb{N}$ we have

$$|e_n(z) - e_m(z)| = \Big|\sum_{k=1}^{n-m} \frac{z^{m+k}}{(m+k)!}\Big| \leq \sum_{k=1}^{n-m} \frac{|z|^{m+k}}{(m+k)!} = e_n(|z|) - e_m(|z|).$$

Since $(e_n(|z|))$ converges (lemma 17.41), the sequence $(e_n(z))$ converges as well. □

19.10 Definition. Let $z \in \mathbb{C}$. We define $\exp(z)$ as follows

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

The function $\exp\colon \mathbb{C} \to \mathbb{C}$ is called the *(complex) exponential function.*

The proof of theorem 17.43 is easily converted into a proof of the next theorem.

19.11 Theorem. *For all $z, w \in \mathbb{C}$ we have* $\exp(z+w) = \exp(z)\exp(w)$. □

19.12 Notation. Here too we will use the notation e^z for $\exp(z)$.

In chapter 17 the notion of continuity for real functions was introduced. This notion is easily extended to the complex case, we give the Heine type of definition of continuity:

19.13 Definition. Let $U \subseteq \mathbb{C}$. A function $f\colon U \to \mathbb{C}$ is called *continuous* in a $\gamma \in U$ if for every sequence (γ_n) in U converging to γ the sequence $(f(\gamma_n))$ converges to $f(\gamma)$. The function f is called continuous if it is continuous in every $\gamma \in U$.

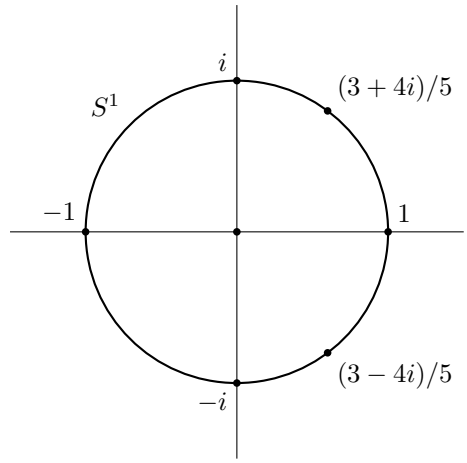


Figure 19.2: The unit circle

As for the real exponential function we have:

19.14 Theorem. *The function* $\exp\colon \mathbb{C} \to \mathbb{C}$ *is continuous.*

Proof. The proof of theorem 17.46 applies here as well. □

Clearly $\overline{e^z} = e^{\overline{z}}$, so by now we have:

$$|e^z|^2 = e^z \cdot \overline{e^z} = e^z \cdot e^{\overline{z}} = e^{z+\overline{z}} = e^{2\Re(z)}.$$

So:

19.15 Proposition. *For all* $z \in \mathbb{C}$ *we have* $|e^z| = e^{\Re(z)}$. □

19.3.2 The unit circle

19.16 Definition. The complex numbers having modulus 1 form the *unit circle* S^1. So:

$$S^1 = \{\, z \in \mathbb{C} \mid |z| = 1 \,\}.$$

If $z_1, z_2 \in S^1$, then also $z_1 z_2 \in S^1$, because $|z_1 z_2| = |z_1| \cdot |z_2| = 1 \cdot 1 = 1$. We also have $z^{-1} \in S^1$ for $z \in S^1$. Clearly the unit circle is a group under multiplication. By proposition 19.15 $|e^{\varphi i}| = 1$ for all $\varphi \in \mathbb{R}$ and so $e^{\varphi i} \in S^1$.

19.17 Proposition. *The map* $\mathbb{R} \to S^1$, $\varphi \mapsto e^{\varphi i}$ *is a homomorphism from the group* $\mathbb{R}$ *(with addition) to the group* S^1 *(with multiplication).*

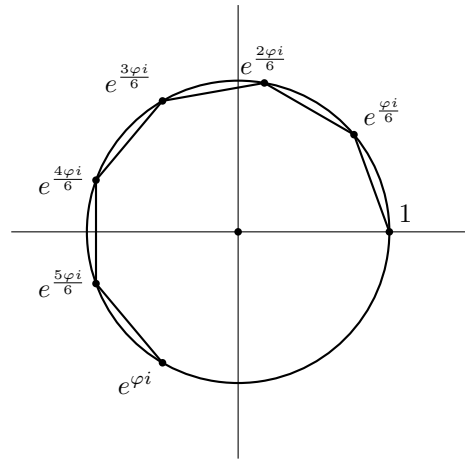


Figure 19.3: Approximation of the arc from 1 to $e^{\varphi i}$

PROOF. For $\varphi, \psi \in \mathbb{R}$ we have $e^{(\varphi+\psi)i} = e^{\varphi i+\psi i} = e^{\varphi i}e^{\psi i}$. □

We take a closer look at this map. Let φ be a positive real number. Then $e^{\varphi i}$ lies on the unit circle. Now the question is: where on the unit circle? For every $m \in \mathbb{N}^+$ we consider the following $m+1$ points on S^1:

$$1, e^{\frac{\varphi i}{m}}, e^{\frac{2\varphi i}{m}}, \ldots, e^{\frac{(m-1)\varphi i}{m}}, e^{\varphi i}.$$

The distances of all pairs of consecutive points are equal:

$$|e^{\frac{(k+1)\varphi i}{m}} - e^{\frac{k\varphi i}{m}}| = |e^{\frac{k\varphi i}{m}}| \cdot |e^{\frac{\varphi i}{m}} - 1| = |e^{\frac{\varphi i}{m}} - 1|,$$

see Figure 19.3. The sum of these distances is $m \cdot |e^{\frac{\varphi i}{m}} - 1|$. We have

$$\lim_m m(e^{\frac{\varphi i}{m}} - 1) = \lim_m \sum_{n=1}^{\infty} \frac{(\varphi i)^n}{m^{n-1} \cdot n!} = \varphi i + \lim_m \sum_{n=2}^{\infty} \frac{(\varphi i)^n}{m^{n-1} \cdot n!}$$

and for $m > 2\varphi$:

$$\Big|\sum_{n=2}^{\infty} \frac{(\varphi i)^n}{m^{n-1} \cdot n!}\Big| < \frac{\varphi^2}{m} \sum_{n=2}^{\infty} \frac{1}{2^{n-2}} = \frac{2\varphi^2}{m}.$$

Hence $\lim_m m(e^{\frac{\varphi i}{m}} - 1) = \varphi i$ and thus $\lim_m m|e^{\frac{\varphi i}{m}} - 1| = \varphi$. The number φ can be interpreted as the length of the arc from 1 to $e^{\varphi i}$, the angle in *radians* of the vector $e^{\varphi i}$ with the positive x-axis. For negative φ, the positive real number $-\varphi$ is the length of the arc from 1 to $e^{\varphi i}$ in the opposite direction. For $\varphi = 2\pi$ we obtain $e^{2\pi i} = 1$. By now we have:

19.18 Theorem. *The map $\mathbb{R} \to S^1$, $\varphi \mapsto e^{\varphi i}$ is a surjective homomorphism from the group $\mathbb{R}$ (with addition) to the group S^1 (with multiplication). Furthermore, for all $\varphi, \psi \in \mathbb{R}$ we have $e^{\varphi i} = e^{\psi i}$ if and only if $\frac{\varphi-\psi}{2\pi} \in \mathbb{Z}$.*

Proof. Proposition 19.17 tells us that it is a homomorphism. If the angle of the vector $z \in S^1$ with the positive x-axis equals φ, then $z = e^{\varphi i}$. So the map is surjective. For $\varphi, \psi \in \mathbb{R}$ the following are equivalent:

$e^{\varphi i} = e^{\psi i}$,

$e^{(\varphi-\psi)i} = 1$,

$\varphi - \psi$ is an integral multiple of 2π. □

19.19 Sine and cosine. In this book the functions sine and cosine are not used. Here only their relation to the complex exponential function is described. Let $\varphi \in \mathbb{R}$. The real part of the complex number $e^{\varphi i}$ is by definition the *cosine* of φ and the imaginary part is the *sine* of φ:

$$e^{\varphi i} = \cos\varphi + \sin\varphi \cdot i.$$

The well-known formulas for the sine and the cosine of a sum are direct consequences:

$$\begin{aligned}\cos(\varphi+\psi) + \sin(\varphi+\psi)i = e^{(\varphi+\psi)i} &= e^{\varphi i}e^{\psi i} \\ &= (\cos\varphi + \sin\varphi\cdot i)(\cos\psi + \sin\psi\cdot i) \\ &= \cos\varphi\cos\psi - \sin\varphi\sin\psi + (\sin\varphi\cos\psi + \cos\varphi\sin\psi)i.\end{aligned}$$

We have: $\cos\varphi = \Re(e^{\varphi i}) = \frac{1}{2}(e^{\varphi i} + e^{-\varphi i})$ and $\sin\varphi = \Im(e^{\varphi i}) = \frac{1}{2i}(e^{\varphi i} - e^{-\varphi i})$. For arbitrary $z \in \mathbb{C}$ we define more generally:

$$\cos z = \tfrac{1}{2}(e^{iz} + e^{-iz}) \quad \text{and} \quad \sin z = \tfrac{1}{2i}(e^{iz} - e^{-iz}).$$

Then

$$\cos z = \sum_{n=0}^{\infty} \frac{(i^n + i^{-n})z^n}{2\cdot n!} = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{(2k)!}$$

and

$$\sin z = \sum_{n=0}^{\infty} \frac{(i^n - i^{-n})z^n}{2i\cdot n!} = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{(2k+1)!}.$$

19.3.3 Roots of unity

See definition 13.44 for the terminology of roots of unity. If ζ is an m-th root of unity of the field $\mathbb{C}$, then $|\zeta|^m = |\zeta^m| = |1| = 1$. So roots of unity lie on the unit circle S^1. The field $\mathbb{C}$ has as many roots of unity as a field might have.

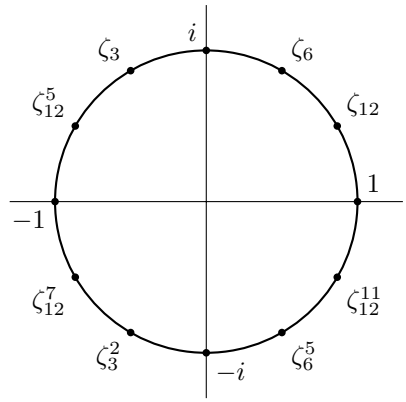


Figure 19.4: The 12-th roots of unity in $\mathbb{C}$

19.20 Lemma. *Let $m \in \mathbb{N}^+$. The number $e^{\frac{2\pi i}{m}}$ is a primitive m-th root of unity of the field $\mathbb{C}$.*

Proof. For all $k \in \mathbb{Z}$ we have $(e^{\frac{2\pi i}{m}})^k = 1 \iff \frac{k}{m} \in \mathbb{Z} \iff m \mid k$. □

19.21 Notation. Let $m \in \mathbb{N}^+$. The number $e^{\frac{2\pi i}{m}}$ is denoted by ζ_m.

The number ζ_m is a kind of standard primitive m-th root of unity. For each $m \in \mathbb{N}$ we thus have a primitive m-th root of unity. A $\zeta \in \mathbb{C}$ is an m-th root of unity if it is a solution of the equation $z^m - 1 = 0$. The number ζ_m is a solution and so are all powers of ζ_m. We have m different solutions: $1, \zeta_m, \zeta_m^2, \ldots, \zeta_m^{m-1}$. Since $z^m - 1 = 0$ is an equation of degree m, there are not more solutions and so:

$$z^m - 1 = (z-1)(z-\zeta_m)(z-\zeta_m^2)\cdots(z-\zeta_m^{m-1}).$$

It follows that the numbers $1, \zeta_m, \zeta_m^2, \ldots, \zeta_m^{m-1}$ are all m-th roots of unity. The primitive m-th roots of unity are the numbers ζ_m^k with $\gcd(k, m) = 1$.

We have seen:

19.22 Proposition. *For every $m \in \mathbb{N}^+$ the number of mth roots of unity in $\mathbb{C}$ equals m. The m-th roots of unity are the numbers $1, \zeta_m, \zeta_m^2, \ldots, \zeta_m^{m-1}$.* □

So the roots of unity in $\mathbb{C}$ are the images of the numbers $2\pi r$ with $r \in \mathbb{Q}$ under the map $\mathbb{R} \to S^1,\ \varphi \mapsto e^{\varphi i}$.

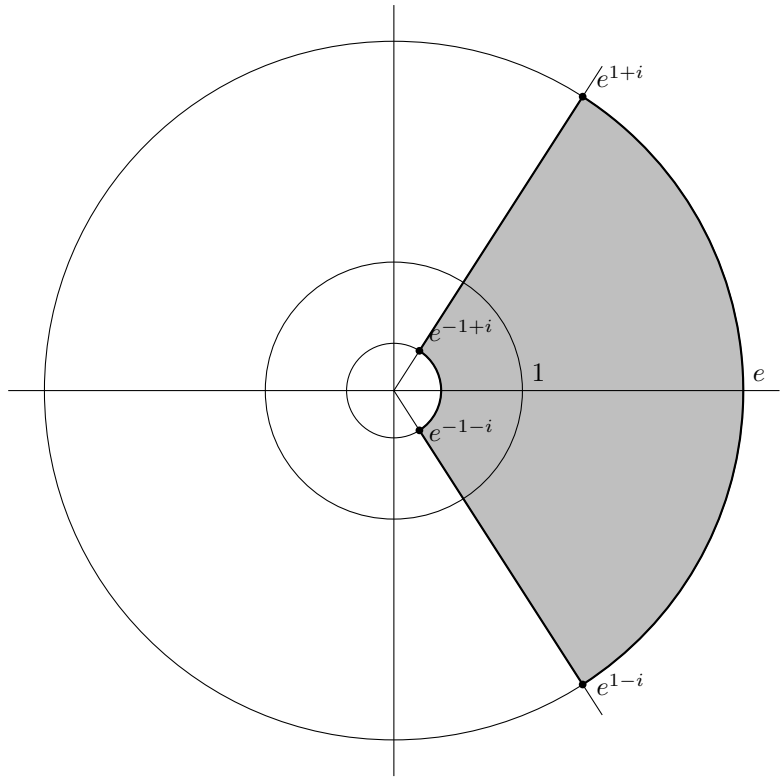


Figure 19.5: The image under exp of the square having the vertices $\pm 1 \pm i$

19.3.4 Complex multiplication

For $z \in \mathbb{C}^*$ we have $z = |z|\frac{z}{|z|}$. The number $\frac{z}{|z|}$ lies on the unit circle and so it can be written as $e^{\varphi i}$ with $\varphi \in \mathbb{R}$. Thus the number z is given by its modulus and the number φ.

19.23 Definition. Let $z \in \mathbb{C}^*$. If $z = |z|e^{\varphi i}$, then φ is called the *argument* of z. The argument is determined up to integral multiples of 2π. If moreover $-\pi < \varphi \leq \pi$, then φ is called the *principal value* of the argument. Notation: $\arg(z) = \varphi$.

Let z be a complex number $\neq 0$ having modulus r and argument φ, and let w be a complex number having modulus s and argument ψ. Then

$$z \cdot w = re^{\varphi i} \cdot se^{\psi i} = rse^{\varphi i + \psi i} = rse^{(\varphi+\psi)i}.$$

The modulus of zw is the product of the moduli of z and w (as we knew already) and

the argument of zw is the sum of the arguments of z and w. So multiplication in $\mathbb{C}$ comes down to the multiplication of the moduli and the addition of the arguments.

By now we have a complete overview of the relation between the additive group $\mathbb{C}$ and the multiplicative group $\mathbb{C}^*$ given by the exponential map $z \mapsto e^z$. It is a surjective group homomorphism $\mathbb{C} \to \mathbb{C}^*$ and we have $e^z = e^w \iff z - w \in 2\pi i\mathbb{Z}$. Figure 19.5 shows the image under the exponential function of the square having the vertices $\pm 1 \pm i$.

19.3.5 m-th roots

The properties of the exponential function imply that every complex number α has an m-th root for any $m \in \mathbb{N}^+$: let $\alpha = e^{a+bi}$, then

$$(e^{\frac{a+bi}{m}})^m = e^{a+bi} = \alpha.$$

Or in terms of modulus and argument: if $\alpha = re^{\varphi i}$ with $r, \varphi i \in \mathbb{R}$ and $r \geq 0$, then

$$(\sqrt[m]{r}e^{\frac{\varphi i}{m}})^m = re^{\varphi i} = \alpha.$$

So for every $\alpha \in \mathbb{C}$ there is a $\beta \in \mathbb{C}$ such that $\beta^m = \alpha$. There are, if $\alpha \neq 0$, m different complex numbers having α as their m-th root:

$$\beta, \zeta_m\beta, \ldots, \zeta_m^{m-1}\beta.$$

Since $z^m - \alpha$ is of degree m, we have:

$$z^m - \alpha = (z - \beta)(z - \zeta_m\beta)\cdots(z - \zeta_m^{m-1}\beta).$$

19.24 Example. We factorize $z^4 - 2$:

$$z^4 - 2 = (z - \sqrt[4]{2})(z - \sqrt[4]{2} \cdot i)(z + \sqrt[4]{2})(z + \sqrt[4]{2} \cdot i).$$

19.25 Example. For $n = 4$ we have $z^4 - 1 = (z-1)(z-i)(z+1)(z+i)$. This factorization can also be found as follows:

$$z^4 - 1 = (z^2 - 1)(z^2 + 1) = (z - 1)(z + 1)(z^2 + 1) = (z - 1)(z + 1)(z - i)(z + i).$$

For $n = 3$: $z^3 - 1 = (z - 1)(z - \zeta_3)(z - \zeta_3^2)$ and

$$z^3 - 1 = (z - 1)(z^2 + z + 1) = (z - 1)(z - (-\tfrac{1}{2} + \tfrac{1}{2}\sqrt{-3})(z - (-\tfrac{1}{2} - \tfrac{1}{2}\sqrt{-3}).$$

We have $\zeta_3 = -\frac{1}{2} + \frac{1}{2}\sqrt{-3}$ and $\zeta_3^2 = -\frac{1}{2} - \frac{1}{2}\sqrt{-3}$.

19.26 Example. We factorize $z^6 - 1$:

$$\begin{aligned} z^6 - 1 &= (z^3 - 1)(z^3 + 1) = (z - 1)(z^2 + z + 1)(z + 1)(z^2 - z + 1) \\ &= (z - 1)(z - \zeta_3)(z - \zeta_3^2)(z + 1)(z + \zeta_3^2)(z + \zeta_3). \end{aligned}$$

Note that $\zeta_6 = -\zeta_3^2$.

19.4 Equations

19.4.1 Quadratic equations

The method of completing the square for the solution of quadratic equations reduces the problem to the extraction of square roots. We have seen that in $\mathbb{C}$ this can be done. So quadratic equations have two solutions, unless the discriminant is 0, then there is one, or as one sometimes says, the two solutions coincide.

19.4.2 Cubic equations

Cardano's method for solving cubic equations involves the extraction of square and cubic roots. If for a given $a \in \mathbb{C}^*$ there is an α such that $\alpha^3 = a$, then the other two cubic roots are $\zeta_3\alpha$ and $\zeta_3^2\alpha$, see subsection 19.3.5.

19.27 Example. In the example of the equation $z^3 - 7z - 6 = 0$ from section 19.1 one solution was found: $z = -2$. The two other solutions are obtained by taking the other two cubic roots. We found $u = \frac{-3+2\sqrt{-3}}{3}$. Another solution is obtained by taking:

$$u = \zeta_3 \cdot \frac{-3+2\sqrt{-3}}{3} = \frac{-1+\sqrt{-3}}{2} \cdot \frac{-3+2\sqrt{-3}}{3} = \frac{-3-5\sqrt{-3}}{6}.$$

To this corresponds $v = \frac{-3+5\sqrt{-3}}{6}$. Thus the solution $z = -1$ is found. The third solution is found by taking

$$u = \zeta_3^2 \cdot \frac{-3+2\sqrt{-3}}{3} = \frac{-1-\sqrt{-3}}{2} \cdot \frac{-3+2\sqrt{-3}}{3} = \frac{9+\sqrt{-3}}{6}$$

and to this corresponds $v = \frac{9-\sqrt{-3}}{6}$. This results in the solution $z = 3$. Alternatively, use that the sum of the solutions equals 0 (= minus the coefficient of z^2), or that their product equals 6 (= minus the constant term). The other two solutions are also easily found using $z^3 - 7z - 6 = (z+2)(z^2 - 2z - 3)$.

> We used that $3^4 + 30\sqrt{-3}$ equals $-3 + 2\sqrt{-3}$ to the power three. This was found by a modulus calculation. The square of the modulus of $3^4 + 30\sqrt{-3}$ equals
>
> $$|3^4 + 30\sqrt{-3}|^2 = 3^8 + 3 \cdot 30^2 = 3^3(3^5 + 10^2) = 3^3 \cdot 343 = 3^3 \cdot 7^3.$$
>
> So if $3^4 + 30\sqrt{-3} = (x + y\sqrt{-3})^3$ with $x, y \in \mathbb{R}$, then $|x + y\sqrt{-3}|^2 = 3 \cdot 7 = 21$, that is $x^2 + 3y^2 = 21$. Solving this equation for $x, y \in \mathbb{Z}$ is easy. A little verification suffices for determining a solution satisfying $3^4 + 30\sqrt{-3} = (x + y\sqrt{-3})^3$. Here $x = -3$ and $y = 2$ were found.

19.4.3 The Fundamental Theorem of Algebra

We have seen that all quadratic and cubic polynomial equations have solutions in $\mathbb{C}$. We will show that all polynomial equations of degree ≥ 1 have solutions. One expresses this by saying that the field $\mathbb{C}$ is *algebraically closed*.

19.28 Fundamental Theorem of Algebra. *Every polynomial equation*

$$z^m + a_1 z^{m-1} + a_2 z^{m-2} + a_3 z^{m-3} + \cdots + a_m = 0$$

with $a_1, \ldots, a_m \in \mathbb{C}$ *and* $m \in \mathbb{N}^+$ *has a solution in* $\mathbb{C}$.

> **Gauß** was the first to give a complete proof. Partial proofs were found by **Laplace**, **Lagrange**, **Argand** and **Euler**. Their proofs rested on algebraic constructions for which only later a solid base was provided.

The proof given here is as elementary as possible. An important ingredient is:

19.29 Proposition. *Polynomial functions on the complex numbers are continuous.*

Proof. As is the case for polynomial functions $\mathbb{R} \to \mathbb{R}$, this is a direct consequences of the rules for limits. □

In the proof of theorem 19.28 we use the notation

$$f(z) = z^m + a_1 z^{m-1} + a_2 z^{m-2} + a_3 z^{m-3} + \cdots + a_m.$$

First some lemmas. In lemma 19.31 it will be shown that $|f(z)|$ reaches a minimal value. For this the continuity of $z \mapsto |f(z)|$ will be used. The proof is completed by showing that the minimal value can not be greater than 0.

19.30 Lemma. *There exists a real number* C *such that* $|f(z)| > \frac{1}{2}|z|^m$ *for all* $z \in \mathbb{C}$ *with* $|z| > C$.

Proof. Take $C = \max(1, 2(|a_1| + \cdots + |a_m|))$. For z with $|z| > C$ we have

$$\begin{aligned} |f(z) - z^m| &= |a_1 z^{m-1} + a_2 z^{m-2} + \cdots + a_m| \\ &\leq |a_1||z|^{m-1} + |a_2||z|^{m-2} + \cdots + |a_m| \\ &\leq (|a_1| + |a_2| + \cdots + |a_m|)|z|^{m-1} < \tfrac{1}{2}|z|^m. \end{aligned}$$

So $|z|^m \leq |f(z)| + |f(z) - z^m| < |f(z)| + \frac{1}{2}|z|^m$, that is $|f(z)| > \frac{1}{2}|z|^m$. □

19.31 Lemma. *There exists a* $\beta \in \mathbb{C}$ *with* $|f(z)| \geq |f(\beta)|$ *for all* $z \in \mathbb{C}$.

Proof. Let C be as given by lemma 19.30. Put $C' = \max(C, \sqrt[m]{2|a_m|})$. Then $|f(z)| > \frac{1}{2}|z|^m \geq \frac{1}{2} \cdot 2|a_m| = |a_m|$ for all $z \in \mathbb{C}$ with $|z| > C'$. So $|f(z)| > |f(0)|$ for all $z \in \mathbb{C}$ with $|z| > C'$. Let D be the disc with center 0 and radius C':

$$D = \{\, z \in \mathbb{C} \mid |z| \leq C' \,\}.$$

For $n \in \mathbb{N}$ let D_n be the set of numbers in D of type $\frac{a+bi}{2^n}$ with $a, b \in \mathbb{Z}$. For every $n \in \mathbb{N}$ the set D_n is a finite nonempty set: $0 \in D_n$ and there are not more than $(2C'+1)^2 \cdot 4^n$ numbers in D_n. Choose for each n an element $\beta_n \in D_n$ with $|f(\beta_n)|$ minimal in $\{\, |f(z)| \mid z \in D_n \,\}$. Write $\beta_n = u_n + v_n i$ with $u_n, v_n \in \mathbb{R}$. The sequence (u_n) in $\mathbb{R}$ is bounded: $|u_n| \leq |\beta_n| < C'$. By theorem 17.21 there is a convergent subsequence $(u_{i(n)})$. The sequence $(v_{i(n)})$ is bounded and again by theorem 17.21 it has a convergent subsequence, say $(v_{j(n)})$. Then $(\beta_{j(n)})$ is a subsequence of (β_n) with $(u_{j(n)})$ and $(v_{j(n)})$ converging. Then $(\beta_{j(n)})$ converges as well. Put $\beta = \lim_n \beta_{j(n)}$. Then $|\beta| = \lim_n |\beta_{j(n)}| \leq C'$ and so $\beta \in D$.

We prove that $|f(z)| \geq |f(\beta)|$ for all $z \in D$. Let $z \in D$. Then there is a sequence (z_n) with $z_n \in D_n$ converging to z. The subsequence $(z_{j(n)})$ converges to z as well. For every n we have $|f(z_{j(n)})| \geq |f(\beta_{j(n)})|$ and so $\lim_n |f(z_{j(n)})| \geq \lim_n |f(\beta_{j(n)})|$, that is $|f(z)| \geq |f(\beta)|$, because the function $z \mapsto |f(z)|$ is continuous.

It remains to prove that $|f(z)| \geq |f(\beta)|$ for all $z \notin D$. For such z we have $|z| > C'$ and so $|f(z)| > |a_n| = |f(0)| \geq |f(\beta)|$. □

Proof of theorem 19.28. From lemma 19.31 it follows that the function $\mathbb{C} \to \mathbb{R}, z \mapsto |f(z)|$ reaches a minimal value: take $\beta \in \mathbb{C}$ such that $|f(z)| \geq |f(\beta)|$ for all $z \in \mathbb{C}$. We will prove that $|f(\beta)| = 0$. Then $f(\beta) = 0$, that is β is a zero of f. We give a proof by contradiction.

> Suppose $|f(\beta)| > 0$. The function $z \mapsto |f(z+\beta)|$ takes a minimal value for $z = 0$. The minimal value of $f(z+\beta)$ is reached for $z = 0$. Dividing by $f(\beta)$ results in a polynomial $g(z) = \frac{f(z+\beta)}{f(\beta)}$, which reaches the minimal value 1 for $z = 0$. Write $g(z)$ as
>
> $$b_0 z^m + b_1 z^{m-1} + \cdots + b_{m-1} z^1 + 1,$$
>
> where $b_0, \ldots, b_{m-1} \in \mathbb{C}$. Let k be such that $b_{m-k+1}, \ldots, b_{m-1} = 0$ and $b_{m-k} \neq 0$. Then
>
> $$g(z) = b_0 z^m + b_1 z^{m-1} + \cdots + b_{m-k} z^k + 1$$
>
> with $b_{m-k} \neq 0$. For every $c \in \mathbb{C}^*$ the function $z \mapsto |g(cz)|$ takes the same values as the function $z \mapsto |g(z)|$ and reaches a minimal value 1 for $z = 0$. Take c such that $c^k = -b_{m-k}$. The coefficient of z^k in $g(cz)$ then equals -1. Put $h(z) = g(cz)$. Then
>
> $$h(z) = c_0 z^m + c_1 z^{c-1} + \cdots + c_{m-k-1} z^{k+1} - z^k + 1$$
>
> where $0 < k \leq m$ and $c_0, \ldots, c_{m-k-1} \in \mathbb{C}$. If $k = m$, then $h(z) = -z^m + 1$ and $h(1) = 0$, contradictory to $|h(z)| \geq 1$ for all z. So $0 < k < m$. We

consider $|h(t)|$ for real numbers t with $0 < t < 1$. If moreover t satisfies $t < \frac{1}{|c_0|+\cdots+|c_{m-k-1}|}$, then

$$\begin{aligned}|h(t)| &\leq |c_0|t^m + |c_1|t^{m-1} + \cdots + |c_{m-k-1}|t^{k+1} + |1-t^k| \\ &\leq (|c_0| + \cdots + |c_{m-k-1}|)t^{k+1} + 1 - t^k < t^k + 1 - t^k = 1.\end{aligned}$$

Contradiction. □

19.32 Corollary. *Let $g(z)$ be a polynomial of degree $n \in \mathbb{N}^+$ with coefficients in $\mathbb{C}$ and leading coefficient α. Then there are $\alpha_1, \ldots \alpha_n \in \mathbb{C}$ with $g(z) = a(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)$.*

Proof. By the Fundamental Theorem of Algebra there is an $\alpha_1 \in \mathbb{C}$ such that $g(\alpha_1) = 0$. In the proof of theorem 9.21 we have seen that $g(z) = (z-\alpha_1)g_1(z)$ with $g_1(z)$ a polynomial of degree $n-1$. Apply the Fundamental Theorem to $g_1(z)$, etc. □

The Fundamental Theorem of Algebra has consequences for the factorization of polynomials over $\mathbb{R}$. If the polynomial $f(z)$ has coefficients in $\mathbb{R}$, then it is also a polynomial over $\mathbb{C}$, since real numbers are only special complex numbers. If $\alpha \in \mathbb{C}$ is a zero of $f(z)$, then $\overline{\alpha}$ is a zero of $f(z)$ as well: $f(\overline{\alpha}) = \overline{f(\alpha)} = \overline{0} = 0$. So under complex conjugation (that is $z \mapsto \overline{z}$) zeros of $f(z)$ map to zeros of $f(z)$. So we can group the n factors in the factorization of $f(z)$:

$$f(z) = (z-\alpha_1)\cdots(z-\alpha_r)(z-\beta_1)(z-\overline{\beta_1})\cdots(z-\beta_s)(z-\overline{\beta_s}),$$

where the zeros $\alpha_1, \ldots, \alpha_r$ are real and the others are not. A product $(z-\beta)(z-\overline{\beta})$ has real coefficients:

$$(z-\beta)(z-\overline{\beta}) = z^2 - (\beta+\overline{\beta})z + \beta\overline{\beta}.$$

So we proved:

19.33 Corollary. *Every polynomial of degree ≥ 1 with coefficients in $\mathbb{R}$ is a product of polynomials of degree ≤ 2 with coefficients in $\mathbb{R}$.* □

19.34 Example. We factorize $x^8 - 1$. The zeros are 1, ζ_8, ζ_8^2 $(= i)$, ζ_8^3 $(= -\overline{\zeta_8})$, ζ_8^4 $(= -1)$, ζ_8^5 $(= -\zeta_8)$, ζ_8^6 $(= -i)$ and ζ_8^7 $(= \overline{\zeta_8})$. We get:

$$\begin{aligned}z^8 - 1 &= (z-1)(z+1)(z-i)(z+i)(z-\zeta_8)(z-\overline{\zeta_8})(z+\zeta_8)(z+\overline{\zeta_8}) \\ &= (z-1)(z+1)(z^2+1)(z^2+\sqrt{2}z+1)(z^2-\sqrt{2}z+1).\end{aligned}$$

Here $(\zeta_8 + \overline{\zeta_8})^2 = i + 2 + (-i) = 2$ is used. Another approach is by factorizing first over $\mathbb{Q}$ as far as possible:

$$z^8 - 1 = (z^4-1)(z^4+1) = (z^2-1)(z^2+1)(z^4+1)$$

Evariste Galois (Bourg la Reine 1811 – Paris 1832)

It was not until 1827 that Galois had his first lessons in mathematics. Mathematics became such an obsession that he neglected other school subjects. He wrote his first article in 1828. It was about continued fractions. During his short life he hardly got any recognition. He was kept in prison from July 14th 1831 until April 29th 1832 because of an alleged threat of king Louis-Philippe. In a duel on May 30th 1832 he got seriously injured and died the next day at the age of twenty. It was the mathematician Liouville who made the publication of Galois's work possible. That was in 1846.

$$
\begin{aligned}
&= (z-1)(z+1)(z^2+1)(z^4+1)\\
&= (z-1)(z+1)(z-i)(z+i)(z^2-i)(z^2+i)\\
&= (z-1)(z+1)(z-i)(z+i)(z-\zeta_8)(z+\zeta_8)(z-\overline{\zeta_8})(z+\overline{\zeta_8}).
\end{aligned}
$$

Every equation of degree n has a solution if $n \geq 1$. For $n = 2$ there is a well-known formula and for $n = 3$ we have Cardano's formula. Cardano's pupil **Ludovico Ferrari** found a formula for $n = 4$. He found a way to reduce an equation of degree 4 to an equation of degree 3. In the beginning of the nineteenth century the Norwegian mathematician **Abel** proved that for equations of degree 5 and higher no general formula (involving only field operations and extraction of roots) can exist. Some years later the French prodigee **Galois** showed that there are concrete polynomial equations having coefficients in $\mathbb{Q}$ for which the solutions are not expressible by roots from rational numbers. An example is the equation $x^5 - 4x + 2 = 0$.

19.5 The Riemann Hypothesis

Here we give a short description of the Riemann Hypothesis. It is one of the big unsolved problems in mathematics.

19.35 Definition. The *Riemann zeta function* is the complex function $s \mapsto \zeta(s)$ with

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Here s is a complex number. For historic reasons one uses for this kind of functions the letter s to denote the complex variable instead of the usual z. Moreover, one

Georg Friedrich Bernhard Riemann (Breselenz 1826 – Selasca 1866)

Bernhard Riemann studied mathematics in Göttingen and later in Berlin, where he learned a lot from Eisenstein, Jacobi and especially from Dirichlet. His thesis on complex functions was remarkable. In it he introduced what are now called Riemann surfaces. Gauß made him return to Göttingen, where he worked on his Habilitation: an extra thesis in Germany which is required for teaching at a university. It took him 30 months and the resulting work has had an enormous impact. He was appointed to professor in Göttingen and became a member of the Berlin Academy of Sciences. To be admitted at the Academy he had to write a report on his recent research. In it he introduced the zeta function as a complex function and formulated what we now call the Riemann hypothesis.

does not write $z = x + yi$, but $s = \sigma + ti$. The complex number n^s is defined as $e^{s \log n}$. By proposition 19.15 we have $|e^{s \log n}| = e^{\Re(s) \log n} = n^{\Re(s)}$, so by theorem 17.26 and exercise 32 of chapter 17 the series converges absolutely for $\Re(s) > 1$. For $s = 1$ it is the diverging harmonic series, see example 16.42. It is possible to extend the function ζ to a neat (= differentiable) function on $\mathbb{C} \setminus \{1\}$. There is a relation (the *functional equation for the Riemann zeta function*) between $\zeta(s)$ and $\zeta(1-s)$:

$$\frac{2^{s-1}\pi^s}{\Gamma(s)} \cdot \zeta(1-s) = \cos\frac{\pi s}{2} \cdot \zeta(s),$$

where $\Gamma(s)$ is an extension of the function $n \mapsto (n-1)!$ defined on $\mathbb{N}^+$. Euler already conjectured this relation. Riemann proved it in 1859. The function value $\zeta(2)$ has been computed by Euler: $\zeta(2) = \frac{\pi^2}{6}$. Euler was even able to compute $\zeta(2n)$ for all $n \in \mathbb{N}^+$:

$$\zeta(2n) = \frac{(-1)^{n-1} 2^{2n-1} B_{2n}}{(2n!)} \pi^{2n},$$

where B_{2n} is the $2n$-th Bernoulli number, see definition 11.33. Moreover, we have $\zeta(0) = -\frac{1}{2}$, $\zeta(-2n) = 0$ and $\zeta(1-2n) = -\frac{B_{2n}}{2n}$ for $n \in \mathbb{N}^+$. Of $\zeta(n)$ for n odd and ≥ 3 not much is known. In 1978 the French mathematician Apéry showed that $\zeta(3)$ is irrational.

The (extended) zeta function has zeros in the even negative integers. Riemann showed that there are infinitely many zeros s satisfying $0 < \sigma < 1$ and that there are no others. The famous *Riemann Hypothesis* is as follows:

$\sigma = \frac{1}{2}$ *for all noninteger zeros of the zeta function.*

The German mathematician **Hans Carl Friedrich von Mangoldt** (1854 – 1925) showed that the Riemann hypothesis is equivalent to the following elementarily formulated conjecture:

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n} = 0,$$

where μ is the Möbius function, see definitions 10.28.

The zeros of the Riemann zeta function give information on the difference of $\pi(x)$ (= number of primes $\leq x$) and $\mathrm{li}(x) = \int_2^x \frac{1}{\log t} dt$, a refinement found by Gauß of the approximation $\frac{x}{\log x}$ of $\pi(x)$.

Exercises

1. Compute: $\frac{1}{1+i}$, $\sqrt{2i}$, $\sqrt[3]{i}$.
2. Solve the following equations:
$$z^2 = i, \quad z^6 = 1, \quad z^3 + z^2 + z + 1 = 0.$$
3. Describe the following subsets of $\mathbb{C}$ geometrically:
$$\{\, 1 + ti \mid t \in \mathbb{R} \,\}, \quad \{\, \tfrac{1}{1+ti} \mid t \in \mathbb{R} \,\}, \quad \{\, z \in \mathbb{C} \mid z + \overline{z} = 2 \,\}.$$
4. The 5-th roots of unity form the vertices of a regular pentagon. A side has length $|1 - \zeta_5|$ and a diagonal $|1 - \zeta_5^2|$. The ratio of these two numbers is the golden ratio. Verify this with a computation in $\mathbb{C}$.
5. Given are $\alpha = \sqrt[5]{\frac{1-i}{1+i}}$ and $\beta = \sqrt[4]{\frac{1-\zeta_5}{1-\zeta_5^{-1}}}$.
   (i) Verify that α and β lie on the unit circle.
   (ii) Are α and β roots of unity?
6. Let $n \in \mathbb{Z}$. Show that the map $S^1 \to S^1$, $z \mapsto z^n$ is a homomorphism from the group S^1 to itself. For which n is this map surjective? For which n injective?
7. Show that the functions $z \mapsto \overline{z}$ and $z \mapsto |z|$ are continuous.
8. Show that the functions $z \mapsto \sin z$ and $z \mapsto \cos z$ are continuous (see page 418). Does it follow that the real functions sin and cos are continuous?
9. Given is the polynomial $f(z) = z^4 + 4z^2 + 2$.
   (i) Determine the zeros of $f(z)$.
   (ii) Let α be one of the zeros. Show that $\alpha + \frac{2}{\alpha}$ is a zero also.

(iii) Let Z be the set of the zeros of $f(z)$. Show that

$$Z \to Z,\ \alpha \mapsto \alpha + \frac{2}{\alpha}$$

is a 4-cycle.

10. Let $m \in \mathbb{N}^+$. Show that there are in $\mathbb{C}$ exactly $\varphi(m)$ primitive roots of unity. ($\varphi(m)$ is the totient of m.)

11. Factorize $z^5 - 1$ as a product of linear and quadratic polynomials with real coefficients.

12. Derive from the Fundamental Theorem of Algebra that a polynomial over $\mathbb{R}$ of odd degree has a zero in $\mathbb{R}$.

13. Let m and n be natural numbers ≥ 1 with $\gcd(m, n) = 1$. Show that there are integers k, l such that $\zeta_{mn} = \zeta_m^k \cdot \zeta_n^l$.

14. Give the modulus and the argument of each of the seven complex solutions of the equation $z^7 - 5 = 0$. The same for the equation $z^7 + 5i = 0$.

15. Solve: $z^5 + \frac{1}{z^5} = 1$. Are the solutions roots of unity?

20 Quadratic Extensions of $\mathbb{Q}$

In section 20.1 we consider the problem:

> For which $a, b \in \mathbb{Q}^*$ are there $x, y \in \mathbb{Q}$ such that $x^2 - ay^2 = b$?

This problem will be solved completely in terms of the prime factorizations of the rational numbers a and b. The outcome is that the equation is solvable in $\mathbb{Q}$ if and only if it is solvable in all completions of $\mathbb{Q}$, so in all fields $\mathbb{Q}_p$ of p-adic numbers and also in the field $\mathbb{R}$. Especially, knowledge of the squares in the p-adic fields will be used. For being solvable in $\mathbb{Q}_p$ we will use Hilbert symbols. These are closely related to the Legendre symbols. The Law of Quadratic Reciprocity and its additional laws come here in the form of a product formula for Hilbert symbols. How all solutions are obtained from just one given solution will be shown first, in section 20.1. It has a geometrical flavor.

20.1 Representation by Quadratic Forms over $\mathbb{Q}$

In chapter 14 we considered the following representation problem:

> Let $a \in \mathbb{Z}$ be not a square. For which $b \in \mathbb{Z}$ are there $x, y \in \mathbb{Z}$ such that $x^2 - ay^2 = b$?

Now we consider a more simple representation problem:

> Let $a \in \mathbb{Q}$ be not a square. For which $b \in \mathbb{Q}^*$ are there $x, y \in \mathbb{Q}$ such that $x^2 - ay^2 = b$?

A formulation symmetric in a and b is given by the following proposition.

20.1 Proposition. *Let a and b be nonzero elements of a field K. Then the following are equivalent:*

(i) *There are $x, y \in K$ such that $x^2 - ay^2 = b$.*
(ii) *There are $x, y \in K$ such that $ax^2 + by^2 = 1$.*
(iii) *There are $x, y \in K$ such that $x^2 - by^2 = a$.*

Proof.

(i)⇒(ii): If $x \neq 0$, then $1 - a(\frac{y}{x})^2 = b(\frac{1}{x})^2$, that is $a(\frac{y}{x})^2 + b(\frac{1}{x})^2 = 1$. If $x = 0$, then $-\frac{b}{a}$ is a square, say $b = -ac^2$ where $c \in K^*$. Then $a(\frac{a+1}{2a})^2 + b(\frac{a-1}{2ac})^2 = 1$.

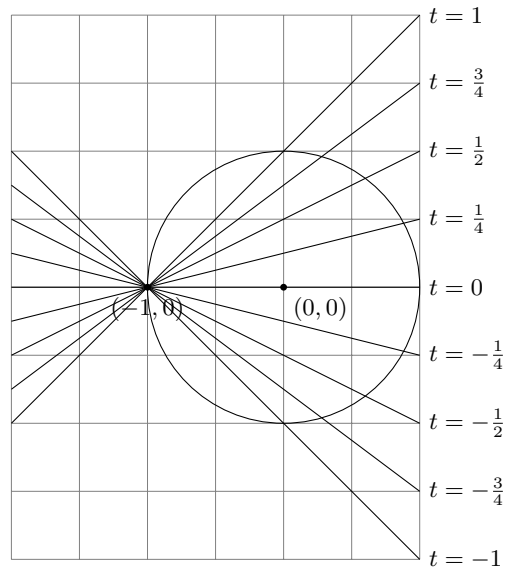


Figure 20.1: Parameterization of $x^2 + y^2 = 1$

(ii)⇒(i): If $y \neq 0$, then $(\frac{1}{y})^2 - a(\frac{x}{y})^2 = b$. If $y = 0$, then a is a square, say $a = c^2$. Then $(\frac{b+1}{2})^2 - a(\frac{b-1}{2c})^2 = b$.

(ii)⇔(iii): This follows from (ii)⇔(i). □

Geometrical meaning

Let a and b be rational numbers $\neq 0$, not both negative. Then the set of points $(x, y) \in \mathbb{R}^2$ with $ax^2 + by^2 = 1$ is an ellipse or a hyperbola. We will show that, if there is a solution in $\mathbb{Q}^2$, there are infinitely many.

The argument is as follows. Let $(u, v) \in \mathbb{Q}^2$ be a solution. Consider the line through (u, v) with slope $t \in \mathbb{Q}$. Its equation is

$$y - v = t(x - u).$$

This line intersects the curve $ax^2 + by^2 = 1$ in two points, one of them being (u, v). Substitution of $y = v + t(x - u)$ in $ax^2 + by^2 = 1$ results in a quadratic equation in x:

$$ax^2 + b(v + t(x - u))^2 = 1.$$

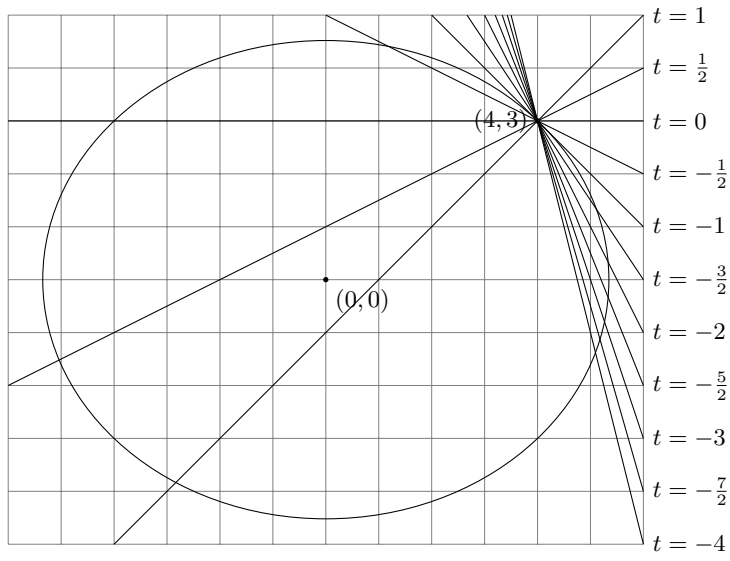


Figure 20.2: Parameterization of $5x^2 + 7y^2 = 143$

Since $x = u$ is a solution, the other solution is an element of $\mathbb{Q}$ as well and the second coordinate of the intersection point is determined by $y = v + t(x - u)$. It also has rational coordinates. Thus for every $t \in \mathbb{Q}$ a point on the curve having rational coordinates is found. Conversely, given a point on the curve having rational coordinates, the line through this point and (u, v) has a rational slope, unless the point is $(u, -v)$, since in that case the line is parallel to the y-axis. It is here also understood that the line tangent to the curve has two coinciding points of intersection.

20.2 Example. The point $(-1, 0)$ lies on the circle $x^2 + y^2 = 1$. Intersect with the line $y = t(x + 1)$. The other point of intersection is:

$$\Big(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\Big),$$

see Figure 20.1. Thus we obtain a parameterization of the rational points on the circle, with the exception of the point $(-1, 0)$. With $t = \frac{a}{b}$ where $a, b \in \mathbb{N}^+$ and $a \leq b$ all Pythagorean triples are found.

20.3 Example. The point $(4, 3)$ lies on the ellipse $5x^2 + 7y^2 = 143$. By the method described above a parameterization of the rational points on the ellipse

(except the point $(4,-3)$) is found:

$$\Big(\frac{28t^2-42t-20}{7t^2+5},\frac{-21t^2-40t+15}{7t^2+5}\Big),$$

see Figure 20.2.

We have seen:

20.4 Theorem. *For a and b nonzero elements of a field K and a not a square: if an equation $ax^2+by^2=1$ has a solution in the field K, then it has more solutions: except for a single solution they can be parameterized by the elements of K.* □

It follows that if, moreover, the field is infinite, so is the number of solutions. Of course the procedure described above leads to a general formula for the solutions. Still the problem remains whether the equation has a solution. For the field $\mathbb{Q}$ this will be solved in section 20.4.

20.2 Adjunction of Square Roots

If an element a of a field K is not a square, then there are ways to extend the field K in such a way that in the larger field the element a is a square. This can be done in a minimal manner, meaning that in the larger field every element is needed. Possibly we already have a larger field in which a is a square. Then inside this field there is a field of the type we are looking for. First an example.

The number 2 is not a square in $\mathbb{Q}$, but it is in the larger field $\mathbb{R}$. If you want the real number $\sqrt{2}$ to be in a larger field, then by addition and multiplication in that field you also have all the $r+s\sqrt{2}$ with $r,s\in\mathbb{Q}$ in this field. These numbers form a subset of $\mathbb{R}$, denoted by $\mathbb{Q}(\sqrt{2})$:

$$\mathbb{Q}(\sqrt{2})=\{\,r+s\sqrt{2}\mid r,s\in\mathbb{Q}\,\}.$$

This subset is closed under addition and multiplication:

$$(r+s\sqrt{2})+(t+u\sqrt{2})=(r+s)+(s+u)\sqrt{2}$$

and

$$(r+s\sqrt{2})\cdot(t+u\sqrt{2})=(rt+2su)+(ru+st)\sqrt{2}.$$

It is also closed under taking inverses: if $r+s\sqrt{2}\neq 0$, then also $r-s\sqrt{2}\neq 0$ and so

$$\frac{1}{r+s\sqrt{2}}=\frac{r-s\sqrt{2}}{r^2-2s^2}=\frac{r}{r^2-2s^2}+\frac{-s}{r^2-2s^2}\sqrt{2}.$$

Since in $\mathbb{R}$ the rules of arithmetic do hold, they are valid in $\mathbb{Q}(\sqrt{2})$ as well. Each of the elements of $\mathbb{Q}(\sqrt{2})$ is determined by an ordered pair (r,s) of rational numbers.

Arithmetic in $\mathbb{Q}(\sqrt{2})$ can be done using a computer, because the elements of $\mathbb{Q}(\sqrt{2})$ can be represented in the computer. In section 21.2 we will make use of this.

The field $\mathbb{Q}$ was extended to $\mathbb{R}$ by analytic means. The field $\mathbb{Q}(\sqrt{2})$ could have been constructed in an algebraic way if $\mathbb{R}$ was not (yet) available: just start with all pairs $(r, s) \in \mathbb{Q}^2$, define addition and multiplication for these pairs by the rules given above and prove that under these operations the set is a field. We will describe this construction more generally.

The construction

Let K be a field with an $a \in K^*$ not being a square in K. We assume that K is *not of characteristic* 2, which means that $1 + 1 \neq 0$, or $2 \neq 0$ as it is usually formulated. We will construct a field $K(\alpha)$ with $\alpha^2 = a$. In such a field we do arithmetic with elements of type $x + y\alpha$ using the identity $\alpha^2 = a$. So we know what we want. It still has to be constructed.

We describe the set $K(\alpha)$ and the operations of addition and multiplication in $K(\alpha)$. We put $K(\alpha) = K \times K = K^2$. The addition is

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$

and the multiplication

$$(x_1, y_1)(x_2, y_2) = (x_1x_2 + ay_1y_2, x_1y_2 + x_2y_1).$$

The rules of arithmetic are easily verified. The zero element is $(0, 0)$ and the unit element is $(1, 0)$. If $(x, y) \in K^2$ differs from $(0, 0)$, then

$$(x, y)(x, -y) = (x^2 - ay^2, 0).$$

Since a is not a square, we have $x^2 - ay^2 \neq 0$. So $(x^2 - ay^2, 0)$ is invertible and hence (x, y) is invertible as well:

$$(x, y)\Big(\frac{x}{x^2 - ay^2}, \frac{-y}{x^2 - ay^2}\Big) = (1, 0).$$

So each nonzero element has an inverse, meaning that $K(\alpha)$ is a field. We have an injective map

$$K \to K(\alpha),\ x \mapsto (x, 0)$$

preserving addition and multiplication. Via this map the field K is isomorphic to $\{\,(x, 0) \mid x \in K\,\}$. We see $K(\alpha)$ as an extension of K: for $(x, 0)$ we write again x and since $(0, 1)^2 = (a, 0)$ we write α for $(0, 1)$. Thus $x + y\alpha$ becomes the notation for (x, y). Arithmetic with the elements $x + y\alpha$ comes down to using the rules of arithmetic together with the identity $\alpha^2 = a$. Often we will denote α by $\sqrt{a}$.

20.5 Terminology. The field $K(\sqrt{a})$ is said to be obtained by *adjunction* of $\sqrt{a}$ to the field K.

Not only the element a of K^* has a square root in $K(\sqrt{a})$:

20.6 Lemma. *The nonsquares of K^* which are squares in $K(\sqrt{a})$ are the elements x^2a with $x \in K^*$.*

Proof. Let $b \in K^*$ be a square in $K(\sqrt{a})$, say $b = (x + y\sqrt{a})^2$. Then $b = x^2 + ay^2 + 2xy\sqrt{a}$ and so $b = x^2 + ay^2$ and $2xy = 0$. Since we assumed that K is not of characteristic 2, it follows that $x = 0$ or $y = 0$. If $y = 0$, then $b = x^2$. If $x = 0$, then $b = ay^2$. □

A classification of fields of type $K(\sqrt{a})$:

20.7 Theorem. *For all nonsquares $a, b \in K^*$ we have*

$$K(\sqrt{a}) \cong K(\sqrt{b}) \iff \text{there exists an } x \in K^* \text{ such that } a = x^2 b.$$

Proof.

$\Rightarrow$ Let $\sigma\colon K(\sqrt{a}) \to K(\sqrt{b})$ be an isomorphism. Then $\sigma(a) \notin K$, since otherwise σ would not be surjective. Since a is a square in $K(\sqrt{a})$, $\sigma(a)$ is a square in $K(\sqrt{b})$. From lemma 20.6 follows that $a = x^2 b$ for an $x \in K^*$.

$\Leftarrow$ Suppose $a = z^2 b$ with $z \in K^*$. Define a map $\sigma\colon K(\sqrt{a}) \to K(\sqrt{b})$ by $\sigma(x + y\sqrt{a}) = x + yz\sqrt{b}$. It is easy to verify that σ is an isomorphism. □

20.8 Definition. The element $\gamma' = x - y\sqrt{a}$ (with $x, y \in K$) is called the *conjugate* in $K(\sqrt{a})$ of $\gamma = x + y\sqrt{a}$. The map $K(\sqrt{a}) \to K(\sqrt{a})$, $\gamma \to \gamma'$ is called *conjugation* in $K(\sqrt{a})$.

20.9 Lemma. *Conjugation in $K(\sqrt{a})$ is an automorphism of $K(\sqrt{a})$. For all $\gamma \in K(\sqrt{a})$ we have $(\gamma')' = \gamma$.*

Proof. For $\gamma_1 = x_1 + y_1\sqrt{a}$ and $\gamma_2 = x_2 + y_2\sqrt{a}$ we have

$$(\gamma_1 + \gamma_2)' = x_1 + x_2 - (y_1 + y_2)\sqrt{a} = x_1 - y_1\sqrt{a} + x_2 - y_2\sqrt{a} = \gamma_1' + \gamma_2',$$

and

$$\begin{aligned}(\gamma_1\gamma_2)' &= (x_1x_2 + ay_1y_2 + (x_1y_2 + y_1x_2)\sqrt{a})' \\ &= x_1x_2 + ay_1y_2 - (x_1y_2 + y_1x_2)\sqrt{a} = (x_1 - y_1\sqrt{a})(x_2 - y_2\sqrt{a}) = \gamma_1'\gamma_2'.\end{aligned}$$

Clearly $(\gamma')' = \gamma$. □

20.10 Definition. The map $N\colon K(\sqrt{a}) \to K$, $x + y\sqrt{a} \mapsto x^2 - ay^2$ is called the *norm* from $K(\sqrt{a})$ to K. In terms of conjugation: $N(\gamma) = \gamma\gamma'$.

20.11 Lemma. *The norm from $K(\sqrt{a})$ to K is multiplicative, that is $N(\gamma_1\gamma_2) = N(\gamma_1)N(\gamma_2)$ for all $\gamma_1, \gamma_2 \in K(\sqrt{a})$.*

Proof. $N(\gamma_1\gamma_2) = \gamma_1\gamma_2(\gamma_1\gamma_2)' = (\gamma_1\gamma_1')(\gamma_2\gamma_2')$. □

The representation problem as given in the previous section can be reformulated using the norm map:

> Let $a \in \mathbb{Q}^*$ be not a square. For which $b \in \mathbb{Q}^*$ is there a $\beta \in \mathbb{Q}(\sqrt{a})$ such that $N(\beta) = b$?

Otherwise put: what is the image of the map $N\colon \mathbb{Q}(\sqrt{a})^* \to \mathbb{Q}^*$? If there are $x, y \in \mathbb{Q}$ such that $x^2 - ay^2 = b$, then this equation has a solution in every field containing $\mathbb{Q}$, such as the completions $\mathbb{R}$ and $\mathbb{Q}_p$ of $\mathbb{Q}$. In section 20.3 we will focus on these completions. After that we will show Hasse's Principle: if there is a solution in each of these completions of $\mathbb{Q}$, then there also is one in $\mathbb{Q}$ itself.

20.12 Definition. Let K and L be fields such that

- $K \subseteq L$,
- Addition and multiplication in K is the restriction of addition and multiplication in L.

Then K is called a *subfield* of L. The field L is called an *extension* of K.

If there exists an $\alpha \in L \setminus K$ such that for each $\beta \in L$ there exist $a, b \in K$ such that $\beta = a + b\alpha$, then L is said to be a *quadratic* extension of K.

The field obtained by adjunction of a square root to a given field K is a quadratic extension of K. If K is not of characteristic 2 all quadratic extensions of K are of this type (exercise 1).

Quadratic extensions of $\mathbb{R}$

In $\mathbb{R}$ all positive numbers are squares, whereas the negative ones are not. So in $\mathbb{R}^*$ there are two types of elements:

$$x^2 \quad \text{and} \quad -x^2$$

with $x \in \mathbb{R}^*$. By theorem 20.7 the construction produces, up to isomorphism, just one new field, namely $\mathbb{R}(\sqrt{-1})$, the field $\mathbb{C}$ (which we studied in chapter 19). We could have started with any nonsquare; the result would have been a field isomorphic to $\mathbb{C}$. It is customary to denote $\sqrt{-1}$ by i. So $\mathbb{C} = \mathbb{R}(i)$.

Quadratic extensions of $\mathbb{F}_p$

In chapter 14 we studied for odd primes p the squares in $\mathbb{F}_p^*$. Half of the elements are squares. Let $\overline{u}$ be a nonsquare. The two types of elements in $\mathbb{F}_p^*$:

$$x^2 \quad \text{and} \quad \overline{u}x^2$$

with $x \in \mathbb{F}_p^*$. Adjunction of a square root of a nonsquare yields a field, unique up to isomorphism. It is denoted by $\mathbb{F}_{p^2}$ and is a field with p^2 elements. Conjugation in $\mathbb{F}_{p^2}$ can also be described as raising to the power p, as we will show now.

Let $u \in \mathbb{Z}$ with $p \nmid u$ and $\left(\frac{u}{p}\right) = -1$, that is $\overline{u}$ is not a square. We have $\mathbb{F}_{p^2} = \mathbb{F}_p(\alpha)$ with $\alpha^2 = \overline{u}$. Since $\binom{p}{k}$ is a multiple of p for all k with $1 \leq k \leq p-1$, we have $(\gamma_1 + \gamma_2)^p = \gamma_1^p + \gamma_2^p$. So for $a, b \in \mathbb{Z}$ we have

$$\begin{aligned}
(\overline{a} + \overline{b}\alpha)^p &= \overline{a}^p + \overline{b}^p\alpha^p = \overline{a} + \overline{b}\alpha^p && \text{(Fermat)} \\
&= \overline{a} + \overline{b}\alpha^{p-1}\alpha = \overline{a} + \overline{b}\,\overline{u}^{\frac{p-1}{2}}\alpha \\
&= \overline{a} - \overline{b}\alpha && \text{(Euler).}
\end{aligned}$$

The field $\mathbb{F}_2$ has no nonsquare. There is a polynomial of degree 2 without zeros in $\mathbb{F}_2$, namely $x^2 + x + 1$. The method of adjunction of a zero of a polynomial $x^2 - a$ can easily be extended to quadratic polynomials having no zeros in the field. In characteristic $\neq 2$ the method of completing the square shows that the result is the same as for the adjunction of the square root of the discriminant of the polynomial. The extension of $\mathbb{F}_2$ obtained by using the polynomial $x^2 + x + 1$ is a field containing an α satisfying $\alpha^2 = \alpha + 1$. The elements of this field are 0, 1, α and $\alpha + 1$. It is a field with four elements and is denoted by $\mathbb{F}_4$. It is a field of characteristic 2, since in this extended field $1 + 1 = 0$ still holds.

Quadratic extensions of $\mathbb{Q}_p$

In subsection 18.5.5 we determined the squares in $\mathbb{Q}_p$. The result is that for odd primes p three extensions of $\mathbb{Q}_p$ can be obtained by adjunction of a square root. Take $u \in \mathbb{Z}$ such that $\left(\frac{u}{p}\right) = -1$. The three fields are

$$\mathbb{Q}_p(\sqrt{p}), \quad \mathbb{Q}_p(\sqrt{u}) \quad \text{and} \quad \mathbb{Q}_p(\sqrt{pu}).$$

On each of these fields a non-Archimedean absolute value is defined by $\gamma \mapsto \sqrt{|N(\gamma)|_p}$. It is an extension of the absolute value on $\mathbb{Q}_p$. These fields are complete with respect to this absolute value. For the first and the last mentioned fields this extension of the absolute value takes new values, for example: from $(\sqrt{p})^2 = p$ follows that $|\sqrt{p}|_p = \frac{1}{\sqrt{p}}$. We will not dwell on this. In section 20.1 we will study the norm to $\mathbb{Q}_p$ on these fields.

For $p = 2$ the situation is somewhat different. There are seven extensions of $\mathbb{Q}_2$ which can be obtained by adjunction of a square root:

$$\mathbb{Q}_2(\sqrt{2}),\ \mathbb{Q}_2(\sqrt{-1}),\ \mathbb{Q}_2(\sqrt{5}),\ \mathbb{Q}(\sqrt{-2}),\ \mathbb{Q}(\sqrt{10}),\ \mathbb{Q}(\sqrt{-5}) \text{ and } \mathbb{Q}_2(\sqrt{-10}).$$

Quadratic extensions of $\mathbb{Q}$

The field $\mathbb{Q}$ is part of the much larger field $\mathbb{R}$. In $\mathbb{R}$ every positive element is a square. The adjunction to $\mathbb{Q}$ of the square root of a positive rational number can be done completely inside $\mathbb{R}$. For the square root of a negative number the larger field $\mathbb{R}(\sqrt{-1}) = \mathbb{C}$ can be used. For $\sqrt{a}$ we take a positive real number if a is positive and $i\sqrt{-a}$ if a is negative. If we mean by $\sqrt{a}$ always a complex number, then $\mathbb{Q}(\sqrt{a}) \cong \mathbb{Q}(\sqrt{b})$ only if $\mathbb{Q}(\sqrt{a}) = \mathbb{Q}(\sqrt{b})$. In chapter 9 we studied which elements of $\mathbb{Q}^*$ are squares: an $x \in \mathbb{Q}^*$ is a square if and only if $x > 0$ and $v_p(x)$ is even for all primes p. For each square free $a \in \mathbb{Z} \setminus \{0\}$ with $a \neq 1$ we have a field $\mathbb{Q}(\sqrt{a})$. By theorem 20.7 there are infinitely many fields of type $\mathbb{Q}(\sqrt{a})$, for every square free $a \neq 1$ there is one.

20.3 Hilbert Symbols

In this section K is one of the completions of $\mathbb{Q}$, that is $K = \mathbb{R}$ or $K = \mathbb{Q}_p$ for some prime p.

20.13 Definition. For $\alpha, \beta \in K^*$ we define

$$(\alpha, \beta) = \begin{cases} 1 & \text{if there are } x, y \in K \text{ such that } \alpha x^2 + \beta y^2 = 1, \\ -1 & \text{otherwise.} \end{cases}$$

The number (α, β) is called the *Hilbert symbol* of α and β. The Hilbert symbol is a map from $K^* \times K^*$ to $\{\pm 1\}$.

> Note that the Hilbert symbol (α, β) only depends on the classes of α and β modulo squares. The field $\mathbb{R}$ has two classes modulo squares, $\mathbb{Q}_p$ has four of these for p odd and $\mathbb{Q}_2$ has eight.

20.14 Proposition. *For all $\alpha, \beta, \gamma \in K^*$*

(i) $(\alpha, \beta) = (\beta, \alpha)$,
(ii) $(\alpha, \beta^2) = 1$
(iii) $(\alpha, -\alpha) = 1$,
(iv) $(\alpha, 1 - \alpha) = 1$ *if* $\alpha \neq 1$,
(v) *if* $(\alpha, \beta) = 1$, *then* $(\alpha, \beta\gamma) = (\alpha, \gamma)$,
(vi) $(\alpha, \alpha) = (\alpha, -1)$.

David Hilbert (Königsberg (now Kaliningrad) 1862 – Göttingen 1943)

Hilbert was an all-round mathematician. In 1900 on the second international congress of mathematicians in Paris he presented 23 problems which he considered to be a challenge for the mathematics of the twentieth century. Among these are the continuum hypothesis (see page 386) and the Riemann hypothesis (see section 19.5). Hilbert made many contributions to number theory, functional analysis (the Hilbert space) and mathematical physics. He participated in the discussions on the foundations of mathematics.

Proof.

(i) By definition the Hilbert symbol is symmetric.
(ii) $\alpha \cdot 0^2 + \beta^2 \cdot (\frac{1}{\beta})^2 = 1$.
(iii) $0^2 - \alpha \cdot 1^2 = -\alpha$ and apply proposition 20.1.
(iv) $\alpha \cdot 1^2 + (1-\alpha) \cdot 1^2 = 1$.
(v) If α is a square, then by (i) and (ii): $(\alpha, \beta\gamma) = 1 = (\alpha, \gamma)$. If α is not a square, then by proposition 20.1 β is the norm of an element of $K(\sqrt{\alpha})$. Then γ is the norm of an element of $K(\sqrt{\alpha})$ if and only if $\beta\gamma$ is such a norm. Then again apply proposition 20.1.
(vi) This follows from (iii) and (v): $(\alpha, \alpha) = (\alpha, (-\alpha)(-1)) = (\alpha, -1)$. □

We will see that $(\alpha, \beta\gamma) = (\alpha, \beta)(\alpha, \gamma)$ for all $\alpha, \beta, \gamma \in K^*$. The parts (ii) and (v) are special cases. This rule says that the Hilbert symbol is multiplicative in the second variable. By (i) it is so in the first variable as well. This is expressed by saying that the Hilbert symbol is *bimultiplicative*.

For each of the completions of $\mathbb{Q}$ we will derive formulas for the Hilbert symbol. Keep in mind that the Hilbert symbol only depends on the classes of the numbers modulo squares. For $K = \mathbb{R}$ the situation is simple:

20.15 Proposition. *For $\alpha, \beta \in \mathbb{R}^*$*

$$(\alpha, \beta) = \begin{cases} -1 & \text{if } \alpha, \beta < 0, \\ 1 & \text{otherwise.} \end{cases}$$

Proof. Clearly, $(-1, -1) = -1$ and $(1, \pm 1) = 1$. □

Next we derive a formula for the Hilbert symbol on $\mathbb{Q}_p$ for p odd. First a lemma.

20.16 Lemma. *There exists a $u \in \mathbb{Z}$ such that both u and $1 - u$ are not squares modulo p.*

Proof. Consider the permutation

$$\mathbb{F}_p \setminus \{\overline{0}, \overline{1}\} \to \mathbb{F}_p \setminus \{\overline{0}, \overline{1}\},\ x \mapsto 1 - x.$$

Not all nonsquares are mapped to a square: in $\mathbb{F}_p \setminus \{\overline{0}, \overline{1}\}$ there are $\frac{p-1}{2}$ nonsquares and $\frac{p-3}{2}$ squares. So there is a nonsquare x such that $1 - x$ is a nonsquare as well. □

Representatives of the classes modulo squares in $\mathbb{Q}_p$ are 1, p, u and pu, where $u \in \mathbb{Z}$ is a nonsquare modulo p. We take u such that both u and $1 - u$ are nonsquares modulo p.

20.17 Theorem. *Let p be an odd prime. Let α and β be elements of $\mathbb{Q}_p^*$. Put $\alpha = p^m\mu$ and $\beta = p^n\nu$ with $m, n \in \mathbb{Z}$ and $\mu, \nu \in \mathbb{Z}_p^*$. Let $\mu \equiv a \pmod p$ and $\nu \equiv b \pmod p$ with $a, b \in \mathbb{Z}$. Then*

$$(\alpha, \beta) = (-1)^{mn\frac{p-1}{2}} \left(\frac{a^n b^m}{p}\right).$$

Proof. For α or β a square the formula is correct. So it suffices to prove the formula for representatives of the classes modulo squares. In all cases the formula turns out to be correct.

a) From lemma 20.16 follows $(u, u) = (u, 1 - u)$ and this equals 1 by proposition 20.14(iv).

b) We will show that $(p, u) = -1$.

   Suppose $(p, u) = 1$. Then there are $x, y \in \mathbb{Q}_p$ such that $px^2 + uy^2 = 1$. Since $|px^2|_p = \frac{1}{p^k}$ with k odd and $|uy^2|_p = \frac{1}{p^l}$ with l even, we have $|px^2|_p \neq |uy^2|_p$ and it follows that $\max(|px^2|_p, |uy^2|_p) = |1|_p = 1$. Since k is odd, this implies that $|px^2|_p < 1$. So $x \in \mathbb{Z}_p$ and $y \in \mathbb{Z}_p^*$. Modulo p we then have $uy^2 \equiv 1$, that is u is a square modulo p. Contradiction.

   So $(p, u) = -1$.

c) From a) and proposition 20.14(v) follows that $(u, pu) = (u, p)$ and this equals -1 by b).

d) By b) we have $(p, v) = -1$ for all $v \in \mathbb{Z}$ with v not a square modulo p, since such v lie in the class modulo squares represented by u. If v is a square modulo p, then v is a square in $\mathbb{Q}_p$. So $(p, v) = \left(\frac{v}{p}\right)$ for all $v \in \mathbb{Z}$ with $p \nmid v$. In particular $(p, -1) = \left(\frac{-1}{p}\right)$. From proposition 20.14(vi) then follows $(p, p) = (p, -1) = \left(\frac{-1}{p}\right)$.

e) Using proposition 20.14(iii) and (v) and also the fact that $(p, v) = \left(\frac{v}{p}\right)$ for all $v \in \mathbb{Z}$ with $p \nmid v$ (proven under d)): $(p, pu) = (p, -u) = \left(\frac{-u}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{u}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{u}{p}\right)$.

f) Using proposition 20.14(vi) and a): $(pu, pu) = (pu, -1) = (p, -1) = \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$. □

A formula for the Hilbert symbol on $\mathbb{Q}_2$:

20.18 Theorem. *Let α and β be elements of $\mathbb{Q}_2^*$. Write $\alpha = 2^m\mu$ and $\beta = 2^n\nu$ with $m, n \in \mathbb{Z}$ and $\mu, \nu \in \mathbb{Z}_2^*$. Let $\mu \equiv a \pmod 8$ and $\nu \equiv b \pmod 8$ with $a, b \in \mathbb{Z}$. Then*

$$(\alpha, \beta) = (-1)^{\frac{a-1}{2}\frac{b-1}{2} + n\frac{a^2-1}{8} + m\frac{b^2-1}{8}}.$$

Proof. It suffices to prove the formula for representatives of the classes modulo squares. We take the following representatives of the seven nontrivial classes:

$$-1, \quad 2, \quad -2, \quad 5, \quad -5, \quad 10, \quad \text{and} \quad -10.$$

So we will compute (α, β) in 28 cases. In each of these cases the result will be as given by the formula.

a) We show that $(-1, -1) = -1$.

> Suppose that $(-1, -1) = 1$. Then there are $x, y \in \mathbb{Q}_2$ such that $-x^2 - y^2 = 1$. By symmetry we can assume that $|x|_2 \geq |y|_2$. Then $1 \leq |x^2|_2$ and so $|x|_2 \geq 1$, say $|x|_2 = 2^k$ with $k \in \mathbb{N}$. Then $2^k x \in \mathbb{Z}_2^*$ and $2^k y \in \mathbb{Z}_2$. Thus we have: $(2^k x)^2 + (2^k y)^2 + 2^{2k} = 0$. In $\mathbb{Z}_2$ squares are congruent to 0 or 1 modulo 4. So each of the three squares is 0 modulo 4. Contradiction, because $(2^k x)^2 \equiv 1 \pmod 4$ since $2^k x \in \mathbb{Z}_2^*$.

So $(-1, -1) = -1$.

b) Some useful cases. We use proposition 20.14 and the fact that the Hilbert symbol depends only on the classes modulo squares.

$$\begin{aligned}
&(-1, 2) = 1,\\
&(-5, -2) = (3, -2) = 1,\\
&(-1, 5) = (-4, 5) = 1.
\end{aligned}$$

c) The remaining cases $(-1, *)$.

$$\begin{aligned}
&(-1, -2) = (-1, 2)(-1, -1) = (-1, -1) = -1\\
&(-1, -5) = (-1, 5)(-1, -1) = (-1, -1) = -1\\
&(-1, 10) = (-1, 2)(-1, 5) = 1\\
&(-1, -10) = (-1, 2)(-1, -5) = -1.
\end{aligned}$$

d) The remaining cases $(2, *)$.

$$\begin{aligned}
&(2, 2) = (2, -1)(2, -2) = 1\\
&(2, -2) = 1\\
&(2, -10) = (2, 10) = (2, 5) = (2, -5) = (-1, -5) = -1.
\end{aligned}$$

e) The remaining cases $(-2, *)$.

$$(-2,-2) = (-2,-1) = -1$$
$$(-2,5) = (2,5) = -1$$
$$(-2,10) = (2,10) = -1$$
$$(-2,-10) = (-2,-5) = 1$$

f) The remaining cases $(\pm 5, *)$.

$$(5,5) = (5,-1) = 1$$
$$(5,-5) = 1$$
$$(5,-10) = (5,10) = (5,2) = -1$$
$$(-5,-5) = (-5,-1) = -1$$
$$(-5,10) = (-5,2) = -1$$
$$(-5,-10) = (-5,-2) = 1$$

g) The remaining cases $(\pm 10, *)$.

$$(10,10) = (10,-1) = 1$$
$$(10,-10) = 1$$
$$(-10,-10) = (-10,-1) = (-5,-1) = -1.$$ □

It is easily checked that the right hand sides of the formulas in the theorems 20.17 and 20.18 are multiplicative in both α and β. So for the Hilbert symbols we have:

20.19 Theorem. *For all* $\alpha, \alpha_1, \alpha_2, \beta, \beta_1, \beta_2 \in K^*$

(i) $(\alpha_1\alpha_2, \beta) = (\alpha_1, \beta)(\alpha_2, \beta)$,
(ii) $(\alpha, \beta_1\beta_2) = (\alpha, \beta_1)(\alpha, \beta_2)$,
(iii) $(\alpha, 1-\alpha) = 1$ *for* $\alpha \neq 1$. □

Note that the other rules are consequences of the three rules in the theorem:

$$(\alpha, \beta^2) = (\alpha, \beta)^2 = 1,$$
$$(\alpha, -\alpha) = \Big(\alpha, \frac{1-\alpha}{1-\frac{1}{\alpha}}\Big) = (\alpha, 1-\alpha)\Big(\alpha, 1-\frac{1}{\alpha}\Big) = \Big(\frac{1}{\alpha}, 1-\frac{1}{\alpha}\Big) = 1,$$
$$(\alpha, \beta) = (\alpha, -\alpha\beta) = (\beta, -\alpha\beta) = (\beta, \alpha).$$

Norms

Let $\alpha \in K^*$ be a nonsquare. Then by proposition 20.1:

$$\beta \in K^* \text{ is the norm of an element of } K(\sqrt{\alpha})^* \iff (\alpha, \beta) = 1.$$

A class modulo squares of K^* either consists of norms only or there is not a single norm in it. The computations of the Hilbert symbols (theorems 20.17 and 20.18) imply that half of the classes modulo squares consist of norms. The trivial class (the class represented by 1) consists of norms: for all $a \in K^*$ we have $N(a) = a^2$.

20.20 Example. Let $\alpha \in \mathbb{R}$ be not a square, that is $\alpha < 0$. Then $\mathbb{R}(\sqrt{\alpha}) = \mathbb{R}(i) = \mathbb{C}$. Modulo squares there are two classes. Only the class of squares consist of norms of elements of $\mathbb{C}^*$.

20.21 Example. The number 2 is not a square in $\mathbb{Q}_5$. In $\mathbb{Q}_5$ there are four classes modulo squares. They are represented by 1, 2, 5 and 10. The class of 1 is the class of squares and these are norms of elements of $\mathbb{Q}_5(\sqrt{2})$. We have $N(\sqrt{2}) = -2$. This is an element of the class of 2, and so this class also consists of norms. In the two other classes there are no norms. In terms of Hilbert symbols: $(1,2) = 1$, $(2,2) = (-2,2) = 1$, $(5,2) = \left(\frac{2}{5}\right) = -1$ and $(10,2) = (5,2)(-1,2) = -1$.

Also 5 is not a square. The norms of elements of $\mathbb{Q}_5(\sqrt{5})^*$ lie in the classes of 1 and 5. We can verify this with Hilbert symbols: $(1,5) = 1$, $(2,5) = \left(\frac{2}{5}\right) = -1$, $(5,5) = (-1,5) = \left(\frac{1}{5}\right) = 1$ and $(10,5) = (5,5)(2,5) = (-1)(-1) = 1$.

20.22 Example. By adjunction of a square root of an element of $\mathbb{Q}_2$ seven different fields can be constructed. In the table below for each of the seven cases it is indicated which of the eight classes modulo squares of $\mathbb{Q}_2$ consist of norms.

	1	−1	2	−2	5	−5	10	−10
$\mathbb{Q}_2(\sqrt{-1})$	+	−	+	−	+	−	+	−
$\mathbb{Q}_2(\sqrt{2})$	+	+	+	+	−	−	−	−
$\mathbb{Q}_2(\sqrt{-2})$	+	−	+	−	−	+	−	+
$\mathbb{Q}_2(\sqrt{5})$	+	+	−	−	+	+	−	−
$\mathbb{Q}_2(\sqrt{-5})$	+	−	−	+	+	−	−	+
$\mathbb{Q}_2(\sqrt{10})$	+	+	−	−	−	−	+	+
$\mathbb{Q}_2(\sqrt{-10})$	+	−	−	+	−	+	+	−

The Product Formula

Numbers $a, b \in \mathbb{Q}^*$ are elements of each of the completions of $\mathbb{Q}$. So for each of these completions we have a Hilbert symbol (a,b).

20.23 Definition. Let a and b be rational numbers $\neq 0$ let p be a prime number. Then we define

$$\left(\frac{a,b}{p}\right) = (a,b),$$

where (a,b) is the Hilbert symbol of $a,b \in \mathbb{Q}_p^*$. Moreover, we define

$$\left(\frac{a,b}{\infty}\right) = (a,b),$$

where (a,b) is the Hilbert symbol of $a,b \in \mathbb{R}^*$. Thus for every p (including ∞) we have a map

$$\mathbb{Q}^* \times \mathbb{Q}^* \to \{\pm 1\}, \quad (a,b) \mapsto \left(\frac{a,b}{p}\right),$$

the *Hilbert symbol* on $\mathbb{Q}$ with respect to p. In this section both prime numbers and ∞ will be called *primes*. We do so only for convenience.

Let $a \in \mathbb{Q}^*$. By proposition 20.1 the representation problem

> For which $b \in \mathbb{Q}^*$ are there $x, y \in \mathbb{Q}$ such that $x^2 - ay^2 = b$?

is equivalent to

> For which $b \in \mathbb{Q}^*$ are there $x, y \in \mathbb{Q}$ such that $ax^2 + by^2 = 1$?

If for a given b there are such x and y in $\mathbb{Q}$, then trivially they also exist in each of the completions of $\mathbb{Q}$, and so $\left(\frac{a,b}{p}\right) = 1$ for all primes p.

20.24 Example. In 14.43 we determined all $n \in \mathbb{Z}$ representable by the form $x^2 - 3y^2$. If an element $b \in \mathbb{Q}^*$ is representable by the form $x^2 - 3y^2$, then $\left(\frac{3,b}{p}\right) = 1$ for all primes p. We will determine all b for which all these Hilbert symbols are trivial. Clearly, $b = -2$ and $b = -3$ are representable. We can assume that b is a square free integer neither divisible by 2 nor by 3, that is $b \equiv \pm 1 \pmod 6$. We have:

- $\left(\frac{3,b}{\infty}\right) = 1$,
- $\left(\frac{3,b}{2}\right) = (-1)^{\frac{b-1}{2}}$, so $b \equiv 1 \pmod 4$ and so $b \equiv 1 \pmod{12}$,
- $\left(\frac{3,b}{3}\right) = \left(\frac{b}{3}\right)$, so $b \equiv 1 \pmod 3$ and so $b \equiv 1, 7 \pmod{12}$,
- for $p \mid b$: $\left(\frac{3,b}{p}\right) = \left(\frac{3}{p}\right)$, so $p \equiv 1, 11 \pmod{12}$.

So we have for square free b with $2, 3 \nmid b$ that b is a product of prime numbers $\equiv \pm 1 \pmod{12}$ and that $b \equiv 1 \pmod{12}$. So the sign of b depends on the parity of the number of prime divisors $\equiv 11 \pmod{12}$. If we also take 2 and 3 in consideration, then we obtain exactly the description given in 14.43. Here we have only shown these conditions to be necessary, not that they are sufficient.

In order for given a and b to compute the Hilbert symbols $\left(\frac{a,b}{p}\right)$, it suffices to compute all but one. This is a consequence of the *product formula* for Hilbert symbols:

20.25 Theorem. *Let a and b be rational numbers $\neq 0$. Then*

$$\prod_p \left(\frac{a,b}{p}\right) = 1,$$

where the product is taken over all primes.

Proof. Since Hilbert symbols are bimultiplicative, it suffices to verify the product formula for both a and b being either a prime number or -1.

a) The product for $a = b = -1$:

$$\left(\frac{-1,-1}{\infty}\right)\left(\frac{-1,-1}{2}\right) = (-1)(-1) = 1.$$

b) For $a = 2$ and $b = -1$ each factor equals 1.
c) For $a = b = 2$ the product is the same as for $a = 2$ and $b = -1$.
d) For $a = p$ an odd prime number and $b = -1$:

$$\left(\frac{p,-1}{2}\right)\left(\frac{p,-1}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{-1}{p}\right).$$

e) For $a = p$ an odd prime number and $b = 2$:

$$\left(\frac{p,2}{2}\right)\left(\frac{p,2}{p}\right) = (-1)^{\frac{p^2-1}{8}}\left(\frac{2}{p}\right).$$

f) For $a = b = p$ an odd prime number the product is the same as for $a = p$ and $b = -1$.
g) For $a = p$ and $b = q$ two different odd prime numbers:

$$\left(\frac{p,q}{2}\right)\left(\frac{p,q}{p}\right)\left(\frac{p,q}{q}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}\left(\frac{q}{p}\right)\left(\frac{p}{q}\right).$$

By the Quadratic Reciprocity Law the product equals 1 in all cases. □

So, conversely, the Quadratic Reciprocity Law and the Supplementary Laws can be retrieved from the product formula for Hilbert symbols.

20.4 Hasse's Principle

In some cases the solvability of equations in $\mathbb{Q}$ is equivalent to their solvability in all completions of $\mathbb{Q}$. In such cases *Hasse's Principle* is said to hold. Hasse's Principle holds for equations $ax^2 + by^2 = 1$:

20.26 Theorem. *For all $a, b \in \mathbb{Q}$ with $a, b \neq 0$:*

$$\text{There are } x, y \in \mathbb{Q} \text{ such that } ax^2 + by^2 = 1 \iff \left(\frac{a,b}{p}\right) = 1 \text{ for all primes } p.$$

Proof. We still have to prove that if $\left(\frac{a,b}{p}\right) = 1$ for all primes p, then there are $x, y \in \mathbb{Q}$ such that $ax^2 + by^2 = 1$. We can assume that a and b are square free integers and that $|a| \leq |b|$. The proof is by induction on $|a| + |b|$.

If $|a| + |b| = 2$, then $a, b = \pm 1$. Since $\left(\frac{-1,-1}{\infty}\right) = -1$ we can assume that $a = 1$ and in this case $ax^2 + by^2 = 1$ is solvable in $\mathbb{Q}$.

Suppose $|a| + |b| > 2$ and $\left(\frac{a,b}{p}\right) = 1$ for all primes p and suppose that $cx^2 + dy^2 = 1$ is solvable in $\mathbb{Q}$ for all square free $c, d \in \mathbb{Z}$ satisfying $|c| + |d| < |a| + |b|$ and $\left(\frac{c,d}{p}\right) = 1$ for all primes p.

Since $|a| + |b| > 2$ and $|a| \leq |b|$, we have $|b| \geq 2$. Let p be a prime divisor of b. If $p \nmid a$ and $p \neq 2$, then $\left(\frac{a,b}{p}\right) = \left(\frac{a}{p}\right) = 1$. So a is a square modulo p, which also is the case if $p \mid a$ or $p = 2$. So a is a square modulo all prime divisors of the square free number b. By the Chinese Remainder Theorem a is a square modulo $|b|$. So there are $c, b' \in \mathbb{Z}$ such that $a = c^2 - bb'$ and $|c| \leq \frac{|b|}{2}$. From $a = c^2 - bb'$ follows that $\left(\frac{a,bb'}{p}\right) = 1$ for all primes p and therefore $\left(\frac{a,b'}{p}\right) = 1$ for all primes p. Furthermore, we have $|bb'| = |c^2 - a| \leq \frac{|b|^2}{4} + |a| \leq \frac{|b|^2}{4} + |b|$ and so $|b'| \leq \frac{|b|+4}{4} < |b|$ (since $|b| \geq 2$). Let b'' be the square free part of b', so $b' = b''d^2$ with b'' square free. Then $|b''| \leq |b'| < |b|$ and $\left(\frac{a,b''}{p}\right) = 1$ for all primes p. By the induction hypothesis there are $x, y \in \mathbb{Q}$ such that $ax^2 + b''y^2 = 1$ and so there also are $x, y \in \mathbb{Q}$ such that $ax^2 + b'y^2 = 1$. We can assume that $a \neq 1$. Then by proposition 20.1 b' is a norm of an element of $\mathbb{Q}(\sqrt{a})$. Also bb' is such a norm: $bb' = c^2 - a = N(c + \sqrt{a})$. So b is a norm of an element of $\mathbb{Q}(\sqrt{a})$, that is there are $x, y \in \mathbb{Q}$ such that $ax^2 + by^2 = 1$. □

> Note that by theorem 20.17 for nonzero $a, b \in \mathbb{Q}$ we have $\left(\frac{a,b}{p}\right) \neq 1$ for only a finite number of odd primes p: if $p \nmid a, b$, then $\left(\frac{a,b}{p}\right) = 1$. So for proving that $ax^2 + bx^2 = 1$ is solvable it suffices to check that $\left(\frac{a,b}{p}\right) = 1$ only for odd primes p dividing ab, for $p = 2$ and for $p = \infty$.

20.27 Example. We determine the numbers $b \in \mathbb{Q}^*$ which are representable by the form $x^2 + 5y^2$. It suffices to consider square free integers. Since 5 is representable, we can assume that $5 \nmid b$. Moreover, clearly only positive b might be representable. We distinguish two cases.

a) $b \in \mathbb{N}^+$ with $2, 5 \nmid b$.
For p a prime divisor of b we have

$$\left(\frac{-5, b}{p}\right) = \left(\frac{-5}{p}\right) = (-1)^{\frac{p-1}{2}} \left(\frac{p}{5}\right) = 1,$$

and also

$$\left(\frac{-5,b}{2}\right) = (-1)^{\frac{b-1}{2}} = 1.$$

So the number b is a product of prime numbers p of two kinds:

I. $p \equiv 1 \pmod 4$ and $p \equiv 1,4 \pmod 5$, that is $p \equiv 1,9 \pmod{20}$,

II. $p \equiv 3 \pmod 4$ and $p \equiv 2,3 \pmod 5$, that is $p \equiv 3,7 \pmod{20}$.

The condition $b \equiv 1 \pmod 4$ means there is an odd number of prime divisors of b which are of kind II.

b) $b \in \mathbb{N}^+$ with $5 \nmid b$ and $2 \mid b$, say $b = 2b'$.
For p an odd prime divisor of b we have

$$\left(\frac{-5,b}{p}\right) = \left(\frac{-5}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{p}{5}\right) = 1,$$

and also

$$\left(\frac{-5,b}{2}\right) = \left(\frac{-5,2b'}{2}\right) = \left(\frac{-5,2}{2}\right)\left(\frac{-5,b'}{2}\right) = -(-1)^{\frac{b'-1}{2}} = 1.$$

So the number b' is a product of prime numbers p of kind I and kind II and the number of prime divisors of kind II is odd.

Because of the product formula there is no need to compute the symbols $\left(\frac{-5,b}{5}\right)$. So we have:

An element $b \in \mathbb{Q}^*$ is representable by the form $x^2 + 5y^2$ if $b > 0$, $v_p(b)$ is even for all prime numbers $p \equiv 11,13,17,19 \pmod{20}$ and moreover, $\sum_{p\in S} v_p(b)$ is even, where S is the set of the prime numbers $p \equiv 3,7 \pmod{20}$ together with the prime number 2.

Exercises

1. Let K be a field of characteristic $\neq 2$ and let L be a quadratic extension of K. Show that L is obtained by adjunction of a square root to K.

2. Show that $\prod_p |a|_p = 1$ for all $a \in \mathbb{Q}^*$. The product is taken over all primes: for p a prime number $|a|_p$ is the p-adic absolute value and $|a|_\infty$ is the ordinary absolute value.

3. Are there rational numbers x and y such that $x^2 - 3\frac{17}{19}y^2 = \frac{5}{2}$?

4. Let K be one of the completions of $\mathbb{Q}$. Let $a \in K^*$, a not a square in K. Let $b, c \in K^*$ be not norms of elements of $K(\sqrt{a})$. Show that bc is a norm of such an element.

5. Let p be a prime number and let $K = \mathbb{F}_p$. Let $c\colon K^* \times K^* \to \{0,1\}$ satisfy

a) $c(a_1a_2, b) = c(a_1, b)c(a_2, b)$ for all $a_1, a_2, b \in K^*$,

b) $c(a, b_1b_2) = c(a, b_1)c(a, b_2)$ for all $a, b_1, b_2 \in K^*$,

c) $c(a, 1-a) = 1$ for all $a \in K^*$ with $a \neq 1$.

Prove that $c(a, b) = 1$ for all $a, b \in \mathbb{F}_p^*$.

21 Quadratic Numbers

In chapter 14 we considered the following representation problem:

> Let $a \in \mathbb{Z}$ be not a square. For which $b \in \mathbb{Z}$ are there $x, y \in \mathbb{Z}$ such that $x^2 - ay^2 = b$?

For only a few a we solved this problem: $a = -1, \pm 2, \pm 3$. In this chapter we focus on some related problems. For b given it is about the Diophantine equation $x^2 - ay^2 = b$ having a solution. Clearly, if a is positive, then the number of solutions is finite. We will see that for a negative there are either no solutions or else infinitely many. This is a consequence of:

> Let $a \in \mathbb{N}$ be not a square. Then the Diophantine equation $x^2 - ay^2 = 1$ has infinitely many solutions.

This will be shown in section 21.3. For a given a we will even give an algorithm for solving Diophantine equations $x^2 - ay^2 = \pm 1$. This algorithm is based on the continued fraction expansion of the irrational number $\sqrt{a}$, which is a solution of a quadratic equation with coefficients in $\mathbb{Q}$. Such numbers are called quadratic. In section 21.2 we study the continued fraction expansions of quadratic numbers. An interesting phenomenon is that quadratic numbers are the numbers with a repeating continued fraction expansion.

21.1 The Discriminant of a Quadratic Number

21.1 Definition. An $\alpha \in \mathbb{C} \setminus \mathbb{Q}$ is called a *quadratic* number if α is a solution of a quadratic equation with coefficients in $\mathbb{Q}$.

Let α be a quadratic number, say α is a solution of $x^2 + px + q = 0$, where $p, q \in \mathbb{Q}$. Write $p = \frac{b}{a}$ and $q = \frac{c}{a}$ with $a \in \mathbb{N}^+$ and $b, c \in \mathbb{Z}$. Then α is a zero of the polynomial $ax^2 + bx + c$, which has coefficients in $\mathbb{Z}$. Divide by $\gcd(a, b, c)$ $(=\gcd(\gcd(a, b), c))$. This results in α being a zero of a polynomial $ax^2 + bx + c$, where $a \in \mathbb{N}^+$, $b, c \in \mathbb{Z}$ and $\gcd(a, b, c) = 1$. To every quadratic number there is a unique such polynomial:

21.2 Proposition. *Let α be a quadratic number. Then there are unique a, b, c such that*

$$a \in \mathbb{N}^+, \quad b, c \in \mathbb{Z}, \quad \gcd(a, b, c) = 1 \quad \textit{and} \quad a\alpha^2 + b\alpha + c = 0.$$

Proof. We still have to show uniqueness. Assume that we also have $a'\alpha^2 + b'\alpha + c' = 0$ with $a' \in \mathbb{N}^+$, $b', c' \in \mathbb{Z}$ and $\gcd(a', b', c') = 1$. Then

$$(a'b - ab')\alpha + (a'c - ac') = 0$$

and so $a'b - ab' = 0$, since otherwise $\alpha \in \mathbb{Q}$. So $a'b = ab'$, and therefore also $a'c = ac'$. Hence $a' \mid \gcd(aa', ab', ac') = a \cdot \gcd(a', b', c') = a$. Similarly $a \mid a'$. And so $a' = a$. But then $b' = b$ and $c' = c$ as well. □

By now we have a correspondence between quadratic numbers and 4-tuples (a, b, c, t) with $a \in \mathbb{N}^+$, $b, c \in \mathbb{Z}$, $t \in \{-1, 1\}$, $\gcd(a, b, c) = 1$ and $b^2 - 4ac$ not a square in $\mathbb{Z}$. Such 4-tuple (a, b, c, t) determines the equation and this equation has two solutions:

$$\alpha = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad \alpha' = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

or for short $\frac{-b+t\sqrt{b^2-4ac}}{2a}$ with $t = \pm 1$. The two solutions are conjugates. Real quadratic numbers correspond to 4-tuples (a, b, c, t) for which, moreover, $b^2 - 4ac > 0$.

21.3 Definition. If the quadratic number α corresponds to the 4-tuple (a, b, c, t), then the number $b^2 - 4ac$ is called the *discriminant* of α, and this number will be denoted by $\operatorname{disc}(\alpha)$. Thus

$$\operatorname{disc}(\alpha) = b^2 - 4ac = a^2 \cdot (\alpha - \alpha')^2.$$

21.4 Lemma. *Let α be a quadratic number. Then $\alpha \pm 1$, $-\alpha$ and $\frac{1}{\alpha}$ are quadratic numbers as well and their discriminant is* $\operatorname{disc}(\alpha)$.

Proof. If α corresponds to (a, b, c, t), then the number $\alpha \pm 1$ corresponds to $(a, \mp 2a + b, a \mp b + c, t)$, $-\alpha$ to $(a, -b, c, -t)$, and $\frac{1}{\alpha}$ to $(\pm c, \pm b, \pm a, \mp t)$. In each case the discriminant is $b^2 - 4ac$. □

21.2 Continued Fraction Expansions of Real Quadratic Numbers

Here we only consider real quadratic numbers. These numbers have (infinite) continued fraction expansions. We will see that the real quadratic numbers are precisely those having a repeating continued fraction expansion. The continued fraction expansion of an $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is closely connected to the course of the transformation

$$\varphi \colon \mathbb{R} \setminus \mathbb{Q} \to \mathbb{R} \setminus \mathbb{Q},\ \alpha \mapsto \frac{1}{\alpha - \lfloor \alpha \rfloor}.$$

The transformation φ can be restricted to a transformation of the set of the real quadratic numbers and even to subsets of real quadratic numbers having a given discriminant:

21.5 Proposition. *Let $\alpha \in \mathbb{R}\setminus\mathbb{Q}$. Then: α is a quadratic number if and only if $\varphi(\alpha)$ is a quadratic number. Moreover, if α is a quadratic number, then* $\operatorname{disc}(\varphi(\alpha)) = \operatorname{disc}(\alpha)$.

Proof. This follows from lemma 21.4 and the definition of the transformation φ. □

Not every $d \in \mathbb{N}^+$ does occur as a discriminant: from $d = b^2 - 4ac$ follows that $d \equiv b^2 \pmod 4$ and so $d \equiv 0, 1 \pmod 4$. Moreover, d is not a square. If these conditions on d are satisfied, then there are real quadratic numbers having discriminant d: if $d \equiv 0 \pmod 4$, then $d = \operatorname{disc}(\frac{1}{2}\sqrt{d})$, and if $d \equiv 1 \pmod 4$, then $d = \operatorname{disc}\big(\frac{1+\sqrt{d}}{2}\big)$.

We will prove that the continued fraction expansion of a real quadratic number repeats. We start with a special case.

21.6 Proposition. *Let α be a real quadratic number satisfying $\alpha > 1$ and $\alpha' < 0$. Then the continued fraction expansion of α repeats.*

Proof. Let d be the discriminant of α. By proposition 21.5 the number $\varphi(\alpha)$ is also a real quadratic number with discriminant d. From

$$\varphi(\alpha) = \frac{1}{\alpha - \lfloor\alpha\rfloor},$$

it follows that

$$\varphi(\alpha)' = \frac{1}{\alpha' - \lfloor\alpha\rfloor} < 0.$$

So the set of real quadratic numbers β with

$$\operatorname{disc}(\beta) = d, \qquad \beta > 1 \qquad \text{and} \qquad \beta' < 0$$

is invariant under φ. This set is finite: for such a β (say it corresponds to the 4-tuple $(a, b, c, 1)$) we have $\frac{c}{a} = \beta\beta' < 0$ and so $c < 0$ and moreover $b^2 + 4a(-c) = d$. So by proposition 21.5 all terms of the sequence $\alpha, \varphi(\alpha), \varphi^2(\alpha), \ldots$ are elements of a finite set. Hence this sequence repeats, and so does the sequence $\lfloor\alpha\rfloor, \lfloor\varphi(\alpha)\rfloor, \lfloor\varphi^2(\alpha)\rfloor, \ldots$. □

21.7 Example. For $\alpha = \sqrt{2}$ we have $\alpha' = -\sqrt{2} < 0$. We already noticed that the continued fraction expansion of $\sqrt{2}$ repeats: $\sqrt{2} = \langle 1, \overline{2}\rangle$.

For a purely repeating continued fraction expansion somewhat stronger conditions are needed:

21.8 Definition. A real quadratic number α is called *reduced* if $\alpha > 1$ and $-1 < \alpha' < 0$.

21.9 Theorem. *A reduced real quadratic number has a purely repeating continued fraction expansion.*

Proof. Let α be a reduced quadratic number with discriminant d. We have

$$\lfloor\alpha\rfloor < \lfloor\alpha\rfloor - \alpha' < \lfloor\alpha\rfloor + 1.$$

Take inverses:

$$\frac{1}{\lfloor\alpha\rfloor + 1} < \frac{1}{\lfloor\alpha\rfloor - \alpha'} < \frac{1}{\lfloor\alpha\rfloor}.$$

So

$$-1 < -\frac{1}{\lfloor\alpha\rfloor} < \varphi(\alpha)' < -\frac{1}{\lfloor\alpha\rfloor + 1} < 0.$$

Hence $\varphi(\alpha)$ is reduced as well. It follows that the set of reduced quadratic numbers with discriminant d is invariant under φ. We will prove that the restriction of φ to this finite set is injective. Suppose β_1 and β_2 are quadratic numbers with discriminant d and $\varphi(\beta_1) = \varphi(\beta_2)$. Then

$$\frac{1}{\beta_1 - \lfloor\beta_1\rfloor} = \frac{1}{\beta_2 - \lfloor\beta_2\rfloor},$$

That is

$$\beta_1 - \lfloor\beta_1\rfloor = \beta_2 - \lfloor\beta_2\rfloor.$$

It follows that

$$\beta_1' - \lfloor\beta_1\rfloor = \beta_2' - \lfloor\beta_2\rfloor$$

and so

$$\lfloor\beta_1\rfloor + (-\beta_1') = \lfloor\beta_2\rfloor + (-\beta_2').$$

Since $0 < -\beta_1', -\beta_2' < 1$ we have therefore

$$\lfloor\beta_1\rfloor = \lfloor\beta_2\rfloor.$$

And so $\beta_1 = \beta_2$ as well. Since the restriction of φ to the set of reduced quadratic numbers having discriminant d is a permutation of this set, $\alpha, \varphi(\alpha), \varphi^2(\alpha), \ldots$ is a purely repeating sequence and so is the continued fraction expansion of α. □

21.10 Example. The number $\sqrt{2}+1$ satisfies $-1 < -\sqrt{2}+1 < 0$ and so $\sqrt{2}+1$ is a reduced real quadratic number. It has a purely repeating continued fraction expansion: $\sqrt{2}+1 = \langle\overline{2}\rangle$.

21.11 Example. We compute the reduced quadratic numbers having discriminant $4 \cdot 34$. Such numbers correspond to 3-tuples (a, b, c) where $a \in \mathbb{N}^+$, $b, c \in \mathbb{Z}$, $\gcd(a, b, c) = 1$, $b^2 - 4ac = 4 \cdot 34$. Clearly, b is even. So we have $(\frac{b}{2})^2 + a(-c) = 34$ with $\frac{b}{2}$ an integer. From $(\frac{b}{2})^2 < 34$ follows that $(\frac{b}{2})^2$ equals 1, 4, 9, 16 or 25. For

a given b we determine all possible numbers a: they are the positive divisors of $34 - (\frac{b}{2})^2$. If α corresponds to (a, b, c), then

$$\alpha = \frac{\sqrt{34} - \frac{b}{2}}{a}.$$

This number is reduced if and only if

$$0 < -\alpha' < 1 < \alpha,$$

that is

$$0 < \sqrt{34} + \frac{b}{2} < a < \sqrt{34} - \frac{b}{2},$$

which is equivalent to

$$0 < 6 + \frac{b}{2} \leq a \leq 5 - \frac{b}{2}.$$

In particular b is negative. We find

$(\frac{b}{2})^2$	$34 - (\frac{b}{2})^2$	a
1	33	
4	30	5, 6
9	25	5
16	18	2, 3, 6, 9
25	9	1, 3, 9

The reduced quadratic numbers having discriminant $4 \cdot 34$ are:

$$\sqrt{34} + 5 = \langle \overline{10, 1, 4, 1} \rangle \qquad \frac{\sqrt{34} + 2}{6} = \langle \overline{1, 3, 3, 1, 1, 1} \rangle$$
$$\frac{\sqrt{34} + 5}{9} = \langle \overline{1, 4, 1, 10} \rangle \qquad \frac{\sqrt{34} + 4}{3} = \langle \overline{3, 3, 1, 1, 1, 1} \rangle$$
$$\frac{\sqrt{34} + 4}{2} = \langle \overline{4, 1, 10, 1} \rangle \qquad \frac{\sqrt{34} + 5}{3} = \langle \overline{3, 1, 1, 1, 1, 3} \rangle$$
$$\frac{\sqrt{34} + 4}{9} = \langle \overline{1, 10, 1, 4} \rangle \qquad \frac{\sqrt{34} + 4}{6} = \langle \overline{1, 1, 1, 1, 3, 3} \rangle$$
$$\frac{\sqrt{34} + 3}{5} = \langle \overline{1, 1, 3, 3, 1, 1} \rangle \qquad \frac{\sqrt{34} + 2}{5} = \langle \overline{1, 1, 1, 3, 3, 1} \rangle$$

The transformation φ restricted to this set has two orbits: one with four elements and one with six. When determining the course of $\sqrt{34}$ under φ (and so its continued fraction expansion) the orbit with four elements is found. So there are six more quadratic numbers having discriminant $4 \cdot 34$. They form another orbit of φ.

Now the general case:

21.12 Theorem. *Let α be a real quadratic number. Then the continued fraction expansion of α repeats.*

Proof. We assume that $\alpha > 1$. (Otherwise replace α by $\varphi(\alpha)$). Next we will show that there is an $n \in \mathbb{N}$ such that $\lfloor \varphi^n(\alpha) \rfloor \neq \lfloor \varphi^n(\alpha)' \rfloor$. It is a proof by contradiction.

Suppose that $\lfloor \varphi^n(\alpha) \rfloor = \lfloor \varphi^n(\alpha)' \rfloor$ for all $n \in \mathbb{N}$.

Let $n \in \mathbb{N}$. We have

$$\alpha = \langle \lfloor \alpha \rfloor, \ldots, \lfloor \varphi^{n-1}(\alpha) \rfloor, \varphi^n(\alpha) \rangle,$$

and so

$$\alpha' = \langle \lfloor \alpha \rfloor, \ldots, \lfloor \varphi^{n-1}(\alpha) \rfloor, \varphi^n(\alpha)' \rangle.$$

We also have

$$\alpha' = \langle \lfloor \alpha' \rfloor, \ldots, \lfloor \varphi^{n-1}(\alpha)' \rfloor, \varphi^n(\alpha') \rangle = \langle \lfloor \alpha \rfloor, \ldots, \lfloor \varphi^{n-1}(\alpha) \rfloor, \varphi^n(\alpha') \rangle.$$

So $\varphi^n(\alpha)' = \varphi^n(\alpha')$, and therefore also $\lfloor \varphi^n(\alpha)' \rfloor = \lfloor \varphi^n(\alpha') \rfloor$.

Hence α and α' have equal continued fraction expansions. This implies that they are equal, which is not the case.

So we may assume that $\lfloor \alpha \rfloor \neq \lfloor \alpha' \rfloor$ and $\alpha > 1$ (take $\varphi^n(\alpha)$ instead of α). From

$$\varphi(\alpha)' = \frac{1}{\alpha' - \lfloor \alpha \rfloor}$$

follows that $\varphi(\alpha)' < 1$. But then $\varphi^2(\alpha)' < 0$ and so by proposition 21.6 the continued fraction expansion of $\varphi^2(\alpha)$ repeats. The quadratic number $\varphi^3(\alpha)$ even has a purely repeating continued fraction expansion, because it is reduced. (Under the assumption $\lfloor \alpha' \rfloor \neq \lfloor \alpha \rfloor \geq 1$). □

The converse holds as well:

21.13 Theorem. *Let α be a real number having a repeating continued fraction expansion. Then α is a quadratic number.*

Proof. Put $\alpha = \langle a_1, \ldots, a_m, \overline{b_1, \ldots, b_n} \rangle$ and $\beta = \langle \overline{b_1, \ldots, b_n} \rangle$. Then $\beta = \varphi^m(\alpha)$ and so it suffices to show that β is a quadratic number. We have:

$$\beta = \langle b_1, \ldots, b_n, \overline{b_1, \ldots, b_n} \rangle = \langle b_1, \ldots, b_n, \beta \rangle.$$

So

$$\beta = \frac{p_n \beta + p_{n-1}}{q_n \beta + q_{n-1}},$$

where $p_i = p_i(b_1 \ldots, b_i)$ and $q_i = q_i(b_1, \ldots, b_i)$. Hence

$$q_n \beta^2 + (q_{n-1} - p_n)\beta - p_{n-1} = 0.$$

So the irrational number β is a quadratic number. □

If α is a reduced real quadratic number, then so is $-\frac{1}{\alpha'}$. There is a simple connection between the continued fraction expansions of these numbers:

21.14 Proposition. *Let* $\alpha = \langle\overline{a_1 \dots, a_n}\rangle$ *with* $a_1, \dots, a_n \in \mathbb{N}^+$. *Then*

$$-\frac{1}{\alpha'} = \langle\overline{a_n, \dots, a_1}\rangle.$$

Proof. Put $\alpha_i = \varphi^{i-1}(\alpha)$. Then

$$\begin{aligned}
\alpha = \alpha_1 &= a_1 + \frac{1}{\alpha_2} && \text{and so} && -\frac{1}{\alpha_2'} &= a_1 + (-\alpha_1')\\
\alpha_2 &= a_2 + \frac{1}{\alpha_3} && && -\frac{1}{\alpha_3'} &= a_2 + (-\alpha_2')\\
&\vdots && && &\vdots\\
\alpha_n &= a_n + \frac{1}{\alpha_{n+1}} = a_n + \frac{1}{\alpha_1} && && -\frac{1}{\alpha_1} &= -\frac{1}{\alpha_{n+1}} = a_n + (-\alpha_n').
\end{aligned}$$

Since $0 < -\alpha_i' < 1$, the proposition follows. □

Python

It is not possible to represent real numbers in a computer: in general an infinite number of data is needed, e.g. for the decimal expansion of a number. However, it is possible to represent quadratic numbers: real quadratic numbers correspond to 4-tuples (a, b, c, t), where $a \in \mathbb{N}^+$, $b, c \in \mathbb{Z}$, $t = \pm 1$, $\gcd(a, b, c) = 1$ and $b^2 - 4ac$ positive and not a square. The functions `ent(alpha)`, `inv(alpha)` and `phi(alpha)` return the floor, the inverse and the φ-image of `alpha`. The function `sub(alpha, n)` returns `alpha` minus the integer `n`. The function `confract(alpha)` returns the continued fraction expansion of `alpha`.

arithmetics.py

```
def ent(alpha):
    return int(divmod(- alpha[1] + alpha[3] * (alpha[1]**2 - 4
    * alpha[0] * alpha[2])**.5, 2 * alpha[0])[0])

def sub(alpha, n):
    return (alpha[0], 2 * alpha[0] * n + alpha[1], alpha[0] * n**2
    + alpha[1] * n + alpha[2], alpha[3])

def inv(alpha):
    s=(-1)**(alpha[2] < 0)
    return (s * alpha[2], s * alpha[1], s * alpha[0], - s
    * alpha[3])
```

arithmetics.py

```
def phi(alpha):
    return inv(sub(alpha, ent(alpha)))

def confract(alpha):
    nrs = []
    exp = []
    while alpha not in nrs:
        a = ent(alpha)
        nrs.append(alpha)
        exp.append(a)
        alpha = inv(sub(alpha, a))
    i = nrs.index(alpha)
    return [exp[:i], exp[i:]]
```

```
>>> ent((1, 0, -37, 1))
6
>>> sub((1, 0, -37, 1), 6)
(1, 12, -1, 1)
>>> inv((1, 12, -1, 1))
(1, -12, -1, 1)
>>> phi((1, 0, -37, 1))
(1, -12, -1, 1)
>>> confract((1, 0, -34, 1))
[[5], [1, 4, 1, 10]]
>>> confract((1, 0, -1141, 1))
[[33], [1, 3, 1, 1, 12, 1, 21, 1, 1, 2, 5, 4, 3, 7, 5, 16, 1, 2, 3,
 1, 1, 1, 2, 1, 2, 1, 4, 1, 8, 1, 4, 1, 2, 1, 2, 1, 1, 1, 3, 2, 1,
16, 5, 7, 3, 4, 5, 2, 1, 1, 21, 1, 12, 1, 1, 3, 1, 66]]
>>> confract((1, -12, -1, 1))
[[], [12]]
```

21.3 Pell's Equation

For $d \in \mathbb{N}^+$, d not a square, we consider *Pell's equation*:

$$x^2 - dy^2 = \pm 1.$$

We will describe a method for obtaining all (infinitely many) solutions $(x, y) \in \mathbb{N}^+ \times \mathbb{N}^+$ of this equation for any given d. The method is based on the continued fraction expansion of the real quadratic number $\sqrt{d}$. We write:

$$\alpha_n = \varphi^{n-1}(\sqrt{d}) \quad \text{and} \quad a_n = \lfloor \alpha_n \rfloor$$

for $n \in \mathbb{N}^+$. Then $\sqrt{d} = \langle a_1, a_2, a_3, \dots \rangle$. For $p_n(a_1, \dots, a_n)$ and $q_n(a_1, \dots, a_n)$ we will simply write p_n and q_n respectively.

21.15 Proposition. *There exists an $n \in \mathbb{N}$ with $n \geq 2$ such that*

$$\sqrt{d} = \langle a_1, \overline{a_2, \ldots, a_n} \rangle \quad \text{and} \quad \begin{cases} a_n = 2a_1 \\ a_{n-i} = a_{i+1} & \text{for } i = 1, \ldots, n-2. \end{cases}$$

Proof. $\sqrt{d}$ is a quadratic number and $(\sqrt{d})' = -\sqrt{d} < -1$. So $\varphi(\sqrt{d}) = \frac{1}{\sqrt{d} - \lfloor \sqrt{d} \rfloor}$ is reduced and therefore has a purely repeating continued fraction expansion, say

$$\frac{1}{\sqrt{d} - a_1} = \langle \overline{a_2, \ldots, a_n} \rangle,$$

It follows that

$$\sqrt{d} = \langle a_1, \overline{a_2, \ldots, a_n} \rangle,$$

or equivalently

$$\sqrt{d} + a_1 = \langle 2a_1, \overline{a_2, \ldots, a_n} \rangle.$$

We have $-\dfrac{1}{(\sqrt{d} + a_1)'} = \dfrac{1}{\sqrt{d} - a_1}$, hence by proposition 21.14

$$\sqrt{d} + a_1 = \langle \overline{a_n, \ldots, a_2} \rangle.$$

This proves the proposition. □

So the proposition says that $a_n = 2a_1$ and that the $(n-2)$-tuple $a_2, \ldots, a_{n-1}$ is symmetric: it coincides with the $(n-2)$-tuple in reverse order.

21.16 Example.

$$\begin{aligned} \sqrt{14} &= 3 + (\sqrt{14} - 3) \\ \frac{1}{\sqrt{14} - 3} &= \frac{\sqrt{14} + 3}{5} = 1 + \frac{\sqrt{14} - 2}{5} \\ \frac{5}{\sqrt{14} - 2} &= \frac{\sqrt{14} + 2}{2} = 2 + \frac{\sqrt{14} - 2}{2} \\ \frac{2}{\sqrt{14} - 2} &= \frac{\sqrt{14} + 2}{5} = 1 + \frac{\sqrt{14} - 3}{5} \\ \frac{5}{\sqrt{14} - 3} &= \sqrt{14} + 3 = 6 + (\sqrt{14} - 3). \end{aligned}$$

So $\sqrt{14} = \langle 3, \overline{1, 2, 1, 6} \rangle$. Indeed, the 3-tuple $1, 2, 1$ is symmetric.

If Pell's equation has a solution, then it can be found in the continued fraction expansion of $\sqrt{d}$:

21.17 Lemma. *Let (p, q) be a solution. Then there exists an $n \in \mathbb{N}^+$ such that $p = p_n$ and $q = q_n$.*

Proof. From $p^2 - dq^2 = \pm 1$ follows $\frac{p^2}{q^2} = d \pm \frac{1}{q^2} \geq 1$ and so $p \geq q$. We then have

$$|p - \sqrt{d} \cdot q| = \frac{1}{p + \sqrt{d} \cdot q} \leq \frac{1}{(1 + \sqrt{d})q} < \frac{1}{2q}$$

and so $|\sqrt{d} - \frac{p}{q}| < \frac{1}{2q^2}$. Now the lemma follows from theorem 17.67. □

Let m be the length of the smallest period of the continued fraction expansion $a_1, a_2, a_3, \ldots$ of $\sqrt{d}$:

$$\sqrt{d} = \langle a_1, \overline{a_2, \ldots, a_{m+1}} \rangle.$$

We can indicate exactly where in the continued fraction expansion of $\sqrt{d}$ solutions can be found:

21.18 Theorem. *Let $n \in \mathbb{N}^+$. We have:*

$$(p_n, q_n) \text{ is a solution} \iff m \mid n.$$

Proof. For all $n \in \mathbb{N}^+$ we have: $\sqrt{d} = \langle a_1, \ldots, a_n, \alpha_n \rangle$ and so

$$\sqrt{d} = \frac{p_n \alpha_{n+1} + p_{n-1}}{q_n \alpha_{n+1} + q_{n-1}}. \tag{21.1}$$

It follows that

$$p_n - q_n \sqrt{d} = \frac{(-1)^n}{q_n \alpha_{n+1} + q_{n-1}}.$$

From $p_n^2 - dq_n^2 = (p_n - q_n\sqrt{d})(p_n + q_n\sqrt{d})$ now follows: if (p_n, q_n) is a solution of $x^2 - dy^2 = 1$, then n is even, and if (p_n, q_n) is a solution of $x^2 - dy^2 = -1$, then n is odd. Moreover we have (also from (21.1)):

$$(p_n - q_n\sqrt{d})\alpha_{n+1} = q_{n-1}\sqrt{d} - p_{n-1}$$

and so

$$(p_n^2 - dq_n^2)\alpha_{n+1} = (q_{n-1}\sqrt{d} - p_{n-1})(p_n + q_n\sqrt{d}) = (-1)^n\sqrt{d} + (\text{integer})$$

$\Rightarrow$: If (p_n, q_n) is a solution, then

$$(-1)^n \alpha_{n+1} = (-1)^n \sqrt{d} + (\text{integer})$$
$$\alpha_{n+1} = \sqrt{d} + (\text{integer}).$$

Then

$$\alpha_{n+2} = \frac{1}{\alpha_{n+1} - \lfloor \alpha_{n+1} \rfloor} = \frac{1}{\sqrt{d} - \lfloor \sqrt{d} \rfloor} = \alpha_2.$$

And so $m \mid n$.

$\Leftarrow$: Suppose $m \mid n$. Then $\sqrt{d} = \langle a_1, \overline{a_2, \ldots, a_{n+1}} \rangle$ and $a_{n+1} = 2a_1$. It follows that

$$\begin{aligned}\sqrt{d} &= \langle a_1, a_2, \ldots, a_{n+1}, \overline{a_2, \ldots, a_{n+1}} \rangle = \langle a_1, a_2, \ldots, a_n, a_1 + \sqrt{d} \rangle \\ &= \frac{p_n(a_1 + \sqrt{d}) + p_{n-1}}{q_n(a_1 + \sqrt{d}) + q_{n-1}}.\end{aligned}$$

So

$$q_n(a_1 + q_{n-1})\sqrt{d} + dq_n = p_n\sqrt{d} + p_n a_1 + p_{n-1},$$

that is

$$\begin{cases} q_n a_1 + q_{n-1} - p_n = 0 \\ p_n a_1 + p_{n-1} - dq_n = 0. \end{cases}$$

Multiply by p_n and q_n respectively:

$$\begin{cases} p_n q_n a_1 + p_n q_{n-1} - p_n^2 = 0 \\ p_n q_n a_1 + p_{n-1} q_n - dq_n^2 = 0. \end{cases}$$

Subtraction yields

$$p_n^2 - dq_n^2 = p_n q_{n-1} - p_{n-1} q_n = (-1)^n. \qquad \square$$

So we have found: the solutions of $x^2 - dy^2 = 1$ are all (p_n, q_n) with $m \mid n$ and n even; the solutions of $x^2 - dy^2 = -1$ are all (p_n, q_n) with $m \mid n$ and n odd. In particular $x^2 - dy^2 = -1$ has solutions only when m is odd.

21.19 Example. $\sqrt{14} = \langle 3, \overline{1, 2, 1, 6} \rangle$. We compute p_4 and q_4:

i:	-1	0	1	2	3	4
a_i:	$-$	$-$	3	1	2	1
p_i:	0	1	3	4	11	15
q_i:	1	0	1	1	3	4

So $15^2 - 14 \cdot 4^2 = 1$. The equation $x^2 - 14y^2 = -1$ has no solutions. The other solutions are (p_8, q_8), (p_{12}, q_{12}), (p_{16}, q_{16}), $\ldots$.

Given a solution a way to find another is as follows: if (x_0, y_0) is a solution, then

$$(x_0 - y_0\sqrt{d})(x_0 + y_0\sqrt{d}) = \pm 1$$

and so also

$$(x_0 - y_0\sqrt{d})^k (x_0 + y_0\sqrt{d})^k = (\pm 1)^k$$

for all $k \in \mathbb{N}^+$. The number $(x_0 + y_0\sqrt{d})^k$ is of type $a + b\sqrt{d}$ with $a, b \in \mathbb{Z}$ and $(x_0 - y_0\sqrt{d})^k$ then equals $a - b\sqrt{d}$. Then (a, b) is a solution as well. In fact: the

solution (x_0, y_0) with the least x_0 is (p_m, q_m) and the other solutions, so (p_{km}, q_{km}) where $k = 2, 3, \dots$, can be obtained by expanding $(x_0 + y_0\sqrt{d})^k$. So from

$$(15 + 4\sqrt{14})^2 = 449 + 120\sqrt{14}$$

it follows that $449^2 - 14 \cdot 120^2 = 1$.

Pell's equation is also encountered when solving other quadratic equations: if for example (x_0, y_0) is a solution of $x^2 - dy^2 = 1$ and (x_1, y_1) a solution of $x^2 - dy^2 = n$, then the number $(x_0 + y_0\sqrt{d})^k(x_1 + y_1\sqrt{d})$ is of type $a + b\sqrt{d}$ where $a, b \in \mathbb{Z}$. Then (a, b) is another solution of $x^2 - dy^2 = n$.

The equation $x^2 - dy^2 = -1$

Whether the equation $x^2 - dy^2 = -1$ has a solution is not for all d directly decidable, but there are some special cases in which this is possible.

21.20 Proposition. *Suppose the Diophantine equation $x^2 - dy^2 = -1$ has a solution. Then d is a sum of two squares in $\mathbb{N}^+$.*

Proof. The period of the continued fraction expansion of $\sqrt{d}$ is odd:

$$\sqrt{d} = \langle a_1, \overline{a_2, \dots, a_k, a_k, \dots, a_2, 2a_1} \rangle.$$

Then for $\beta = \varphi^k(\sqrt{d})$ we have:

$$\beta = \langle \overline{a_k, \dots, a_2, 2a_1, a_2, \dots, a_k} \rangle.$$

Furthermore, $\operatorname{disc}(\beta) = \operatorname{disc}(\sqrt{d}) = 4d$. Let β belong to the quadruple $(a, b, c, 1)$. From the symmetry in the continued fraction expansion of β follows

$$-\frac{1}{\beta'} = \beta,$$

that is $\beta\beta' = -1$. So $\frac{c}{a} = -1$, that is $c = -a$. So $4d = b^2 - 4ac = (2a)^2 + b^2$, and therefore $d = a^2 + (\frac{b}{2})^2$, where $\frac{b}{2} \in \mathbb{Z}$ since b is even. □

21.21 Example. $\sqrt{13} = \langle \overline{3, 1, 1, 1, 1, 6} \rangle$, $\varphi^3(\sqrt{13}) = \frac{\sqrt{13}+2}{3}$, and so $13 = 2^2 + 3^2$.

The converse does not hold: 34 is a sum of two squares

$$34 = 3^2 + 5^2,$$

but the period of the continued fraction expansion of $\sqrt{34}$ is of even length:

$$\sqrt{34} = \langle 5, \overline{1, 4, 1, 10} \rangle.$$

Example 21.11 shows that there are two reduced quadratic numbers having discriminant $4 \cdot 34$ which have the symmetry in their continued fraction expansion we are looking for: $\langle \overline{3, 1, 1, 1, 1, 3} \rangle$ and $\langle \overline{1, 1, 3, 3, 1, 1} \rangle$. The identity $34 = 3^2 + 5^2$ can be rewritten as $(\frac{5}{3})^2 - 34(\frac{1}{3})^2 = -1$. The equation $x^2 - 34y^2 = -1$ has no integer solution, but it has a rational one.

In special cases the converse does hold:

21.22 Theorem. *Let p be a prime number such that $p \equiv 1 \pmod 4$. Then the Diophantine equation $x^2 - py^2 = -1$ has a solution.*

Proof. Let m be the length of the shortest period of the continued fraction expansion of $\sqrt{p}$.

Suppose m is even. Then (p_m, q_m) is a solution of $x^2 - py^2 = 1$, that is

$$(p_m - 1)(p_m + 1) = p_m^2 - 1 = pq_m^2.$$

Then p_m is odd and q_m is even. So

$$\frac{p_m - 1}{2} \cdot \frac{p_m + 1}{2} = p\Big(\frac{q_m}{2}\Big)^2,$$

and $\frac{p_m-1}{2}, \frac{p_m+1}{2}, \frac{q_m}{2} \in \mathbb{N}^+$. Clearly $\gcd(\frac{p_m-1}{2}, \frac{p_m+1}{2}) = 1$. So:

$$\begin{cases} \dfrac{p_m - 1}{2} = u^2 \\ \dfrac{p_m + 1}{2} = pv^2 \end{cases} \quad \text{or} \quad \begin{cases} \dfrac{p_m - 1}{2} = pv^2 \\ \dfrac{p_m + 1}{2} = u^2 \end{cases}$$

for some $u, v \in \mathbb{N}^+$. Then

$$u^2 - pv^2 = \pm\Big(\frac{p_m + 1}{2} - \frac{p_m - 1}{2}\Big) = \pm 1,$$

while $u < p_m$. Contradiction.

So m is odd, that is (p_m, q_m) is a solution of $x^2 - py^2 = -1$. □

Again we obtain theorem 14.8:

21.23 Corollary. *Let p be a prime number congruent to 1 modulo 4. Then p is a sum of two squares.* □

The Indian mathematician **Brahmagupta** (Ujjain(?) 598 – 670) was the first to study Pell's equation systematically. The algorithm for the solution originates in the works of **Bhaskara** (Vijayapura 1114 – Ujjain 1185). In the seventeenth century progression was made by **Fermat**, the English **John Wallis** (1616–1703) and the Irish **William Brouncker** (1620–1684). **Euler** gave a full solution. The formulation in terms of continued fractions goes back to the Italian-born French mathematician **Joseph-Louis Lagrange** (1736 – 1813). The naming of the equation is from Euler, but probably is a result of a misunderstanding, since as far as known, not a single text by John Pell mentions the equation, it is unlikely that he contributed to its solution.

Python

For the computation of the continued fraction expansion of $\sqrt{d}$ it is not necessary to keep track of its course since for repetition it suffices to look at the floor of the quadratic number: it has to be twice the floor of $\lfloor\sqrt{d}\rfloor$, see also exercise 15. The function `pell(d)` returns a triple (x, y, a), where (x, y) is the solution of $x^2 - dy^2 = (-1)^a$ and a the length of the period of the continued fraction expansion of $\sqrt{d}$.

arithmetics.py

```
def pell(d):
  alpha = (1, 0, -d, 1)
  e = a = ent(alpha)
  p, q, r, s = 0, 1, 1, 0
  i = 0
  b = 2 * e
  while a != b:
      alpha = inv(sub(alpha, a))
      p, q, r, s = r, s, a * r + p, a * s + q
      a = ent(alpha)
      i = i + 1
  return (r, s, i)
```

```
>>> pell(34)
(35, 6, 4)
>>> pell(1141)
(10367823941572239636237125215, 30693385322765657197397208, 58)
>>> pell(94)
(2143295, 221064, 16)
>>> pell(95)
(39, 4, 4)
>>> pell(1234567)
(20371567825887579087969922203933358793493846332810110697412723191699811071244735562
4, 18334417735362515888338401277540899079617602694996527932674628391416461414984172
5, 124)
>>> opl = pell(123456789)
>>> len(str(opl[0])), len(str(opl[1])), opl[2]
(4197, 4193, 8164)
>>> opl = pell(2**41 - 1)
>>> len(str(opl[0])), len(str(opl[1])), opl[2]
(316673, 316667, 615482)
```

Exercises

1. Let $m \in \mathbb{N}^+$, m not a square. Determine the discriminant of $\sqrt{m}$ and also of $\frac{1+\sqrt{m}}{2}$.
2. The numbers $\sqrt{2}$ and $\sqrt{3}$ are quadratic. Is $\sqrt{2} + \sqrt{3}$ quadratic too?

3. Determine all reduced quadratic numbers with discriminant 60. Also give their continued fraction expansions.

4. Determine all reduced quadratic numbers with discriminant 80. Also give their continued fraction expansion.

5. Let $\alpha = \langle a_1, \overline{a_2, \ldots, a_n} \rangle$ with $a_1, \ldots, a_n \in \mathbb{N}^+$. Suppose that $a_n = 2a_1$ and $a_{n-i} = a_{i+1}$ for $i = 1, \ldots, n-2$. Prove that $\alpha^2 \in \mathbb{Q}$.

6. Let d be a multiple of 4 plus 3. Show that the Diophantine equation $x^2 - dy^2 = -1$ has no solution. Does it have a solution in $\mathbb{Q}$?

7. Give a solution of the the Diophantine equation $x^2 - 29y^2 = -1$ and also of the Diophantine equation $x^2 - 29y^2 = 1$.

8. Is the Diophantine equation $x^2 - 33y^2 = -1$ solvable?.

9. Are there rational numbers x and y such that $x^2 - 34y^2 = 15$? Are there integral solutions? How many?

10. Let $d \in \mathbb{N}^+$, d not a square. Suppose that the Diophantine equation $x^2 - dy^2 = 2$ is solvable. Is it possible to find solutions using a continued fraction?

11. Take in exercise 1 of chapter 17 $\alpha = 2$. Then $a = 2$ and

$$\begin{cases} a_0 = 2 \\ a_{n+1} = \frac{1}{2}(a_n + \frac{2}{a_n}) \quad \text{for all } n \in \mathbb{N}. \end{cases}$$

The sequence (a_n) converges to $\sqrt{2}$. Show that

$$a_n = \frac{p_{2^n}}{q_{2^n}}$$

for all $n \in \mathbb{N}$. Here $\frac{p_m}{q_m}$ is the m-th convergent of the continued fraction expansion of $\sqrt{2}$.

12. (i) Determine the continued fraction expansion of $\sqrt{13}$.
    (ii) Determine the least $y \in \mathbb{N}^+$ for which $13y^2 - 1$ is a square. Also determine the least but one y having this property.
    (iii) How many reduced quadratic numbers are there in the set

$$A = \{ \varphi^n(\sqrt{13}) \mid n \in \mathbb{N} \}?$$

    (iv) Are there reduced quadratic numbers α such that $\mathrm{disc}(\alpha) = \mathrm{disc}(\sqrt{13})$ and $\alpha \notin A$?

13. (i) Find natural numbers x, y such that $x^2 - 29y^2 = \pm 1$.
    (ii) Determine all natural numbers x, y such that $x^2 - 25y^2 = \pm 1$.

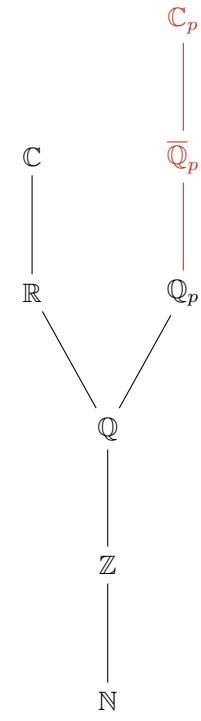
14. Let $\alpha = \langle \overline{1, 1, 2, 1} \rangle$.
    (i) Determine the continued fraction expansion of $3 + \alpha$ and also of $\frac{1}{\alpha - 1}$.

(ii) Compute α.

15. Prove that for each nonsquare $d \in \mathbb{N}^+$ there is a unique reduced real quadratic number γ such that $\operatorname{disc}(\gamma) = 4d$ and $\lfloor \gamma \rfloor = 2 \cdot \lfloor \sqrt{d} \rfloor$.

Number Systems

We have built number systems starting with Peano's axioms for the natural numbers and nothing more than that. In the diagram the successive extensions are indicated. The field $\mathbb{R}$ is obtained as a completion of $\mathbb{Q}$. By adjoining a square root of -1 we get the field $\mathbb{C}$ in which all polynomials of positive degree have roots: $\mathbb{C}$ is algebraically closed. The fields $\mathbb{Q}_p$ are other completions of $\mathbb{Q}$. To achieve that all polynomials of positive degree have roots infinitely many roots of polynomials have to be adjoined. We have not done so in this book. The field one obtains this way is denoted by $\overline{\mathbb{Q}}_p$. It has in a natural way an absolute value and is complete with respect to this absolute value. However, this field is not algebraically closed. Again one can adjoin roots of polynomials. This finally results in a field called $\mathbb{C}_p$ which is both complete and algebraically closed. It is a long way from $\mathbb{N}$ to $\mathbb{C}_p$. In between $\mathbb{Q}$ and $\mathbb{C}$, as well as in between $\mathbb{Q}$ and $\mathbb{C}_p$ there are infinitely many other fields, e.g. the quadratic extensions we studied in Part V.



Notations

$a \in A$	a is an element of the set A	8
$a \notin A$	a is not an element of the set A	8
$A \subseteq B$	the set A is a subset of the set B	10
$\{\, a \in A \mid P(a) \,\}$	the set of all $a \in A$ such that $P(a)$	11
$\emptyset$	the empty set	11
$A \cap B$	the intersection of the sets A and B	11
$A \cup B$	the union of the sets A and B	12
$A \setminus B$	the difference of the sets A and B	12
B^c, B'	the complement of the set B (within a given set A)	12
$\neg p$	not p	14
$p \wedge q$	p and q	14
$p \vee q$	p or q	14
$p \Rightarrow q$	if p, then q	14
$p \iff q$	p if and only if q	14
$\mathcal{P}(A)$	the power set of the set A	15
$A \div B$	the symmetric difference of the sets A and B	17
$\mathbb{N}$	the set of natural numbers	30
$\mathbb{N}^+$	the set of natural numbers $\neq 0$	30
$f\colon A \to B$, $A \xrightarrow{f} B$	the map f from set A to set B	57
$f\colon a \mapsto b$	the map f maps the element a to the element b	57
(a, b)	an ordered pair	59
$A \times B$	the (Cartesian) product of the sets A and B	59
$\Gamma(f)$	the graph of the map f	59
$f_*(U)$, $f(U)$	the image under the map f of the subset U of the domain of f	61
$f^*(V)$, $f^{-1}(V)$	the inverse image under the map f of the subset V of the codomain of f	61
f^{-1}	the inverse of the bijection f	62
1_A	the identity transformation of the set A	62
$\tau_{a,b}$	the transposition of the elements a and b	62

gf, $g \circ f$	the composition of the maps f and g	63
$A \approx B$	the sets A and B are equipotent	66
$\underline{n}$	for the natural number the set of the natural numbers $1, 2, \ldots, n$	66
$\mathbb{N}_n$	for the natural number n the set of the natural numbers $0, 1, \ldots, n-1$	66
$\#(A)$	the number of elements of a finite set A	68
B^A	the set of all maps from the set A to the set B	71
χ_U	the characteristic function of the subset U	72
$D(f)$	the support of the $\{0,1\}$-valued function f	72
(a_n)	the sequence $a_0, a_1, a_2, \ldots$	76
$\mathcal{R}(A)$	the set of sequences in the set A	77
$\mathcal{F}(A)$	the set of finite sequences in the set A	77
$\mathcal{F}_n(A)$	the set of finite sequences of length n in the set A	77
$n!$	n factorial	78
f^n	the n-th iterate of the transformation f	81
$[a]_f$	the f-class of a in the domain of the map f	90
A_f	the set of f-classes in the domain A of f	90
$[a]_\Phi$	the class of the partition Φ containing the element a	91
$[a]_\Phi$	the class of the partition Φ represented by a	91
$[a]_\sim$	the equivalence class of the element a with respect to the equivalence relation $\sim$	94
$A/\sim$	the set of equivalence classes with respect to the equivalence relation $\sim$ in the set A	94
$\mathbb{Z}$	the ring of integers	96
$a = \min(U)$	the least element of the ordered set U	106
$a = \max(U)$	the greatest element of the ordered set U	106
$q_b(a)$	the quotient of the integer a divided by the integer $b > 0$	116
$r_b(a)$	the remainder of the integer a after division by the integer $b > 0$	116
$\sum_{k=0}^{n-1} a_k$	the sum of the elements a_0, a_1, ... and a_{n-1}	117
$\sum_{i \in I} a(i)$	the sum of the elements $a(i)$ over all $i \in I$	117
$[x_{n-1}, \ldots, x_1, x_0]_g$	the number $x_0 g^0 + x_1 g^1 + \cdots + x_{n-1} g^{n-1}$	122
$\mathcal{R}_c(A)$	the set of sequences in the set A having a c-tail	122
$\mathbb{Q}$	the field of rational numbers	136
$\lfloor x \rfloor$	the floor (or entier) of x	142
$\gcd(a, b)$	the greatest common divisor of the integers a and b	147
$\mathrm{lcm}(a, b)$	the least common multiple of the integers a and b	155

$\langle x_1, \ldots, x_n \rangle$	the continued fraction of length n	157
$v_p(a)$	the p-adic value of a	174
$f * g$	the Dirichlet product of the arithmetical functions f and g	183
μ	the Möbius function	185
φ	Eulers's totient function	188
$\mathrm{Inj}(A.B)$	the set of injective maps from the set A to the set B	193
$\mathcal{P}_k(A)$	the set of all subsets of A with k elements	195
$\binom{n}{k}$	the binomial coefficient n over k	195
$\left\langle {n \atop k} \right\rangle$	the number of admitted words in $\{\mathtt{0}, \mathtt{1}\}$ of length n with exactly k ones	204
c_n	the n-th Catalan number	204
B_n	the n-th Bernoulli number	216
$B_n(x)$	the n-th Bernoulli polynomial	218
$\left\{ {n \atop k} \right\}$	the Stirling number of the second kind	224
$[a]_\sigma$	the orbit of a under the permutation σ	231
$(a_1 \quad a_2 \quad \cdots \quad a_n)$	the n cycle of $a_1, \ldots, a_n$	232
$D(\sigma)$	the support of the permutation σ	233
$\left[{n \atop k} \right]$	the Stirling number of the first kind	236
$\mathrm{sgn}(\sigma)$	the sign of the permutation σ	238
$a \equiv b \pmod{m}$	a is congruent to b modulo m	247
$[a]_m$	the residue class of a modulo m	247
$\mathrm{o}_m(a)$	the least $k \in \mathbb{N}^+$ with $\overline{a}^k = \overline{1} \in \mathbb{Z}/m$	256
$\mathbb{Z}/m$	the ring of residue classes modulo m	247
$\left(\frac{a}{p}\right)$	the Legendre symbol (a and b integers)	274
$\left(\frac{a}{b}\right)$	the Jacobi symbol (a and b rational)	281
$\pi(n)$	the number of primes $\leq n$	297
F_n	the n-th Fermat number	308
$\lim_n a_n$	the limit of the sequence (a_n) for $n \to \infty$	331
$\lim_n^{(p)} a_n$	the p-adic limit of the sequence (a_n)	345
$\mathbb{R}$	the field of the real numbers	354
$\exp(x)$	the (real) exponential function	369
$\log(x)$	the (natural) logarithm of the real x	371
$\langle a_1, a_2, a_3, \ldots \rangle$	the infinite continued fraction of $a_1, a_2, a_3, \ldots$	376
$A \preceq B$	there exists an injective map from A to B	383
$A \prec B$	there exists an injective map from A to B and there is no surjective map from A to B	383
$\mathbb{Q}_p$	the field of the p-adic numbers	391

$\mathbb{Z}_p$	from chapter 18 onwards: the ring of p-adic integers	394
$\mathbb{Z}_p^{(n)}$	set of all $\alpha \in \mathbb{Z}_p$ with $\alpha \equiv 1 \pmod{p}$	397
$K(\sqrt{a})$	the field obtained by adjoining $\sqrt{a}$ to the field K	435
(α, β)	the Hilbert symbol on K, where K is one of the completions of $\mathbb{Q}$	439
$\left(\frac{a,b}{p}\right)$	the Hilbert symbol on $\mathbb{Q}$ with respect to the prime p (including $p = \infty$)	444
$\mathbb{C}$	the field of the complex numbers	412
$\Re(z)$	the real part of the complex number z	413
$\Im(z)$	the imaginary part of the complex number z	413
S^1	the unit circle in the complex plane	416
ζ_m	the complex root of unity $e^{\frac{2\pi i}{m}}$	419
$\zeta(s)$	the Riemann zeta function	426

Index

Learning mathematics just by doing it the way mathematicians do

This is a textbook for beginning mathematics students. Knowledge of school mathematics is not presumed. Mathematics is presented from scratch with the construction of the number system from the natural numbers via the rationals and the reals to the complex numbers as a common thread. For the interested reader also the other possible completions of the rationals – the p-adic numbers – are constructed. Applications in combinatorics, number theory and cryptography are given. With many examples and exercises.

ISBN 978-94-9329-682-4

Radboud University

www.radbouduniversitypress.nl