DEBRECENI EGYETEM

Number theory and its applications

Szerző:
Dr. Hajdu Lajos

A tanulmány elkészítését a „A Debreceni Egyetem fejlesztése a felsőfokú oktatás minőségének és hozzáférhetőségének együttes javítása érdekében” az **EFOP-3.4.3-16-2016-00021** számú projekt támogatta. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Number theory and its applications

Lajos Hajdu

Contents

Introduction

These lecture notes are composed mostly for the mathematics students of the University of Debrecen. Its main purpose is to help their studies of number theory and its applications, and to deepen their knowledge - and to point out possible applications. From this perspective, the most important parts of the book are those concerning prime tests, factorization, and the LLL algorithm. At the same time we put an emphasize on the fundamentals of these chapters, that is, we intend to describe the most important underlying notions (such as for instance prime numbers, lattices, etc.). This includes the precise establishment of the basic concepts and their essential properties. As an example, we mention the chain algebraic number theory - Diophantine approximation - continued fractions - LLL algorithm. The most important applications of the LLL algorithm for us, can be interpreted as higher generation variants of algorithms using continued fractions, so studying continued fractions is inevitable for the introduction and use of the LLL algorithm. However, understanding continued fractions is not possible without discussing Diophantine approximation - and that cannot be done without tools from algebraic number theory.

Beside these, better understanding is helped with certain chapters included mostly for the sake of 'interest'. For example, the chapter concerning Pythagorean triples helps us to understand better the essence of prime factorization, or treating Waring's problem, to prove some of our statements we need to apply tools from the theory of the Legendre symbol and lattice theory - and these tools are of utmost importance for certain 'applied' types of problems, as well. As a further example, we can mention the chapter concerning arithmetic functions, too. Here, on the one hand, we discuss highly important theorems and assertions (e.g. concerning Euler's φ function), but on the other hand we mention things which in themselves are not important for our later studies. However, without them our insight on and understanding

of the topic would be much poorer and weaker.

We also mention that in making our lecture notes, we used the sources mentioned in the References at several points. Studying these sources (as well as other related books) can be useful for the Reader, too.

We close this brief introduction by saying that we find the problems and topic discussed in the present book rather colorful and interesting. In many cases it is worth and exciting to think over the arising questions and to try to find extensions, generalizations of the presented results. So we hope that the Reader will not only find the lecture notes helpful and useful, but also a source of fun!

Debrecen, December 2019

Lajos Hajdu

Chapter 1

Arithmetic functions

In this chapter we study arithmetic functions. First we formulate general theorems. Then we study some particular arithmetic functions separately. We put a special emphasize on Euler's φ function and its properties, and for interest, also on the σ function. Beside this we shall closely investigate two transformations on the set of arithmetic functions. We conclude the chapter with studying the behavior of the average of arithmetic functions.

1.1 Additive and multiplicative arithmetic functions

In this section we investigate certain relevant properties of arithmetic functions. First we introduce the notion of arithmetic functions.

Definition 1.1.1 *A function defined on the set of natural numbers, mapping into the set of complex numbers, is called an arithmetic function.*

Now we introduce two important classes of arithmetic functions.

Definition 1.1.2 *An arithmetic function f is called additive, if for all $a, b \in \mathbb{N}$, $\gcd(a, b) = 1$ we have*

$$f(ab) = f(a)f(b). \tag{1.1}$$

If (1.1) is valid for all $a, b \in \mathbb{N}$, then f is totally additive.

Definition 1.1.3 *An arithmetic function f is multiplicative, if for all $a, b \in \mathbb{N}$, $\gcd(a, b) = 1$*

$$f(ab) = f(a)f(b) \tag{1.2}$$

holds. If (1.2) is valid for all $a, b \in \mathbb{N}$, then f is totally multiplicative.

Example 1.1.1 Some examples for the above notions:

- the arithmetic function $f(n) = \lg(n)$, $f : \mathbb{N} \to \mathbb{R}$ is totally additive,
- the arithmetic function $g(n) = n$, $g : \mathbb{N} \to \mathbb{N}$ is totally multiplicative,
- the identically zero function is totally multiplicative and totally additive at the same time.

The following theorems show that an additive or a multiplicative arithmetic function can take only special values at 1.

Theorem 1.1.1 *Let f be an additive arithmetic function. Then we have $f(1) = 0$.*

Proof. Since f is additive,

$$f(1) = f(1 \cdot 1) = f(1) + f(1)$$

holds. Hence $f(1) = 0$, which was to be proved. □

Theorem 1.1.2 *Let f be a not identically zero multiplicative arithmetic function. Then $f(1) = 1$ holds.*

Proof. As f is multiplicative, we have

$$f(1) = f(1 \cdot 1) = f(1) \cdot f(1).$$

After rearrangement this gives

$$f(1)(f(1) - 1) = 0,$$

so $f(1) = 1$ or $f(1) = 0$. If $f(1) = 0$ would hold, then again by the multiplicativity of f, for arbitrary $n \in \mathbb{N}$, as $\gcd(n, 1) = 1$ we would get

$$f(n) = f(1 \cdot n) = f(1) \cdot f(n) = 0,$$

implying $f(n) = 0$. However, then f is the identically zero function, which is excluded. Hence we necessarily have $f(1) = 1$, which proves our claim. $\square$

Remark 1.1.1 The above theorems imply that an additive or multiplicative arithmetic function can take only the values $0, 1$ at 1. This immediately implies that additivity and multiplicativity is very rare: if, for example for an arithmetic function f we have $f(1) = 2$, then this f can be neither additive nor multiplicative, independently of the values of f taken at the other natural numbers.

Our next theorems show that it is sufficient to know the values of additive and multiplicative functions on prime powers (furthermore, in the 'total' case on primes).

Theorem 1.1.3 *Let f be an additive arithmetic function. Then f is uniquely determined by its values taken on the prime powers. Further, if f is totally additive, then f is uniquely determined already by its values taken at the primes.*

Proof. Since f is an additive function, we have $f(1) = 0$. If $n \in \mathbb{N}$, $n > 1$, then by the fundamental theorem of arithmetic we have

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}.$$

(Here and later on the above expression is the prime factorization of the number n in question: $p_1, \dots, p_r$ are different primes, and the exponents $\alpha_1, \dots, \alpha_r$ are non-negative integers.) As f is additive, by induction on r we obtain

$$f(n) = f(p_1^{\alpha_1}) + \dots + f(p_r^{\alpha_r}).$$

If moreover, f is totally additive, then

$$f(n) = \alpha_1 f(p_1) + \dots + \alpha_r f(p_r)$$

also holds. This proves our statement. $\square$

Theorem 1.1.4 *Let f be a not identically zero multiplicative arithmetic function. Then f is uniquely determined by its values taken on the prime powers. Further, if f is totally multiplicative, then f is uniquely determined already by its values taken at the primes.*

Proof. As f is a not identically zero multiplicative function, we have $f(1) = 1$. If $n \in \mathbb{N}$, $n > 1$, then by the fundamental theorem of arithmetic n can be written as

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}.$$

Thus by induction on r, by the multiplicativity of f we easily get that

$$f(n) = f(p_1^{\alpha_1}) \dots f(p_r^{\alpha_r}).$$

Furthermore, if f is totally multiplicative, then we also have

$$f(n) = f(p_1)^{\alpha_1} \dots f(p_r)^{\alpha_r}.$$

Hence our theorem is proved. □

Remark 1.1.2 We point out a similarity of additive and multiplicative functions with linear mappings. Namely, it is sufficient to know a linear mapping (between vector spaces V_1 and V_2) on a basis of V_1, from this the linear mapping can be extended to the whole V_1. The 'building bricks' of vector spaces are the elements of a basis, while those of the set of natural numbers are the primes (prime powers) - this is the root of the analogy.

At this point we formulate two further observations.

Remark 1.1.3 Consider the arithmetic function

$$\delta(n) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{otherwise.} \end{cases}$$

One can easily see that $\delta(n)$ is totally multiplicative. It is also clear that δ and the identically zero (totally multiplicative) arithmetic function takes the same values on the prime powers. Thus the condition in the previous theorem, excluding the identically zero function, is necessary.

The following assertions are so-called structure theorems: they show that the additive and multiplicative arithmetic functions, for the appropriate operations form certain structures.

Our first statement of this type concerns additive and totally additive arithmetic functions.

Theorem 1.1.5 *The additive arithmetic functions form an Abelian group with respect to addition. The totally additive arithmetic functions also form an Abelian group with respect to addition.*

Proof. Let f and g be additive arithmetic functions. Then for any $a, b \in \mathbb{N}$, $\gcd(a, b) = 1$ we have

$$\begin{aligned}(f+g)(ab) = f(ab) + g(ab) = (f(a) + f(b)) + (g(a) + g(b)) = \\ = (f(a) + g(a)) + (f(b) + g(b)) = (f+g)(a) + (f+g)(b),\end{aligned}$$

that is, $f + g$ is also additive. So the set of additive arithmetic functions is closed with respect to addition.

Let $O(n)$ be the identically zero arithmetic function. Obviously, O is additive. Further, for any additive arithmetic function f and $n \in \mathbb{N}$ we have

$$(f + O)(n) = f(n) + O(n) = f(n) = O(n) + f(n) = (O + f)(n),$$

so O is the zero element of the structure.

Let now f be an arbitrary additive arithmetic function. By the usual notation

$$(-f)(n) = -f(n)$$

we obtain that for any $n \in \mathbb{N}$

$$\begin{aligned}(f + (-f))(n) = f(n) - f(n) = 0 = O(n) = \\ = 0 = -f(n) + f(n) = ((-f) + f)(n)\end{aligned}$$

is valid. That is, any additive arithmetic function has an additive inverse. In other words: our structure is closed with respect to taking additive inverse.

Let now f, g be additive arithmetic functions. Then for any $n \in \mathbb{N}$

$$(f + g)(n) = f(n) + g(n) = g(n) + f(n) = (g + f)(n)$$

holds. This shows that addition is a commutative operation on the set of additive arithmetic functions.

Finally, let f, g, h be arithmetic functions. Then for any $n \in \mathbb{N}$ we have

$$((f+g)+h)(n) = (f(n)+g(n))+h(n) = f(n)+(g(n)+h(n)) = (f+(g+h))(n).$$

Hence addition is associative on the structure.

Summarizing the above assertions, we see that the set of additive arithmetic functions forms an Abelian group with respect to addition indeed.

The statement concerning the totally additive functions can be proved similarly, so it is left to the Reader. □

We continue our investigations with structures formed by multiplicative functions. As we shall see, these structures are less rich. The reason for this is that if a function f assumes zero, then it cannot have a multiplicative inverse. We give our statements without proofs: they could be proved similarly to Theorem 1.1.5.

Theorem 1.1.6 *The multiplicative arithmetic functions form a commutative semigroup with unity with respect to multiplication. The totally multiplicative arithmetic functions also form a commutative semigroup with unity with respect to multiplication.*

The structure of the set of (totally) multiplicative arithmetic functions which do not take zero value is already similar to the ones formed by additive arithmetic functions. The proof of this statement, as it is similar to the proof of Theorem 1.1.5, is omitted again.

Theorem 1.1.7 *The nowhere zero multiplicative arithmetic functions form an Abelian group with respect to multiplication. The nowhere zero totally multiplicative arithmetic functions also form an Abelian group with respect to multiplication.*

1.2 Some particular arithmetic functions

In this section we consider certain particular arithmetic functions and study some of their important properties.

Definition 1.2.1 *Let $n \in \mathbb{N}$. Introduce the following arithmetic functions:*

- $\chi(n)$ *is the number of different prime divisors of* n,
- $\nu(n)$ *is the number of prime factors of* n *(counted with multiplicity),*
- $d(n)$ *is the number of divisors of* n,
- $\sigma(n)$ *is the sum of divisors of* n,
- $\varphi(n)$ *is the number of positive integers coprime to* n*, not greater than* n *(so* φ *is Euler's function),*

- $\mu(n)$ *is the Möbius function, that is*

$$\mu(n) = \begin{cases} 1, & \text{if } n = 1, \\ (-1)^r, & \text{if } n \text{ is of the shape } n = p_1 \dots p_r, \\ 0, & \text{otherwise.} \end{cases}$$

In what follows, we show that the arithmetic functions introduced above, hold some of the previously defined properties. We start with the additive functions, then we turn to the multiplicative ones.

Theorem 1.2.1 *The function $\chi(n)$ is additive.*

Proof. Let $a, b \in \mathbb{N}$, $\gcd(a, b) = 1$. Assume first that one of a, b is 1. By symmetry we may suppose that $a = 1$. Then clearly

$$\chi(ab) = \chi(b) = 0 + \chi(b) = \chi(a) + \chi(b)$$

holds. Let now $a > 1$, $b > 1$, with prime factorizations

$$a = p_1^{\alpha_1} \dots p_r^{\alpha_r}, \quad b = q_1^{\beta_1} \dots q_t^{\beta_t}.$$

As $\gcd(a, b) = 1$, the primes p_i $(i = 1, \dots, r)$ and q_j $(j = 1, \dots, t)$ are pairwise coprime. Hence we get

$$\chi(ab) = r + t = \chi(a) + \chi(b),$$

which proves the statement. □

Remark 1.2.1 It is easy to check that χ is not totally additive.

Theorem 1.2.2 *The function $\nu(n)$ is totally additive.*

Proof. Let $a, b \in \mathbb{N}$. If one of a, b equals 1, then by symmetry again we may assume that $a = 1$. Then certainly

$$\nu(ab) = \nu(b) = 0 + \nu(b) = \nu(a) + \nu(b).$$

On the other hand, if $a > 1$, $b > 1$ with prime factorizations

$$a = p_1^{\alpha_1} \dots p_r^{\alpha_r}, \quad b = q_1^{\beta_1} \dots q_t^{\beta_t}$$

then we obtain

$$\nu(ab) = \alpha_1 + \cdots + \alpha_r + \beta_1 + \cdots + \beta_t = \nu(a) + \nu(b),$$

which proves our statement. □

Now we turn to multiplicative functions.

Theorem 1.2.3 *The functions $d(n)$, $\sigma(n)$, $\varphi(n)$, $\mu(n)$ are multiplicative.*

Proof. First we prove that $d(n)$ and $\sigma(n)$ are multiplicative. Let $a, b \in \mathbb{N}$, $\gcd(a, b) = 1$. If $d \mid ab$, then the coprimality of a and b yields that there exist $a', b' \in \mathbb{N}$, such that $a' \mid a$, $b' \mid b$ and $a'b' = d$. Thus if $a_1, \ldots, a_k$ and $b_1, \ldots, b_\ell$ are the divisors of a and b, respectively, then the divisors of ab are

$$a_1b_1, \ldots, a_1b_\ell, a_2b_1, \ldots, a_2b_\ell, \ldots, a_kb_1, \ldots, a_kb_\ell.$$

Hence

$$d(ab) = k\ell = d(a)d(b),$$

and also

$$\sigma(ab) = \sum_{i=1}^{k}\sum_{j=1}^{\ell} a_ib_j = \sum_{i=1}^{k} a_i \sum_{j=1}^{\ell} b_j = \sigma(a)\sigma(b).$$

Now we show that $\mu(n)$ is multiplicative. For this, let $a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$. If a or b is 1, then using symmetry we may assume that $a = 1$. Then

$$\mu(ab) = \mu(b) = 1 \cdot \mu(b) = \mu(a)\mu(b).$$

Thus we may suppose that $a > 1$ and $b > 1$. Let the prime factorizations of a and b be given by

$$a = p_1^{\alpha_1} \ldots p_r^{\alpha_r} \quad \text{and} \quad b = q_1^{\beta_1} \ldots q_t^{\beta_t}.$$

If here any of the exponents α_i $(i = 1, \ldots, r)$ or β_j $(j = 1, \ldots, t)$ is greater than 1, then on the one hand

$$\mu(ab) = 0,$$

and on the other hand

$$\mu(a) = 0 \quad \text{or} \quad \mu(b) = 0.$$

Thus in any case

$$\mu(ab) = \mu(a)\mu(b)$$

follows. Hence we may suppose that

$$\alpha_1 = \cdots = \alpha_r = \beta_1 = \cdots = \beta_t = 1.$$

However, then

$$\mu(ab) = (-1)^{r+t} = (-1)^r(-1)^t = \mu(a)\mu(b),$$

so $\mu(n)$ is multiplicative, as well.

Finally, we prove that φ is multiplicative. Let again $a, b \in \mathbb{N}$, $\gcd(a, b) = 1$. If $a = 1$ or $b = 1$, then assuming by symmetry that $a = 1$, we obtain

$$\varphi(ab) = \varphi(b) = 1 \cdot \varphi(b) = \varphi(a)\varphi(b).$$

Otherwise, let

$$r_1, \ldots, r_{\varphi(a)}$$

be a reduced residue system modulo a, with $1 = r_1 < \ldots < r_{\varphi(a)} < a$. Then the integers between 1 an ab, being coprime to a are just the elements of the table

$$\begin{array}{ccc} r_1 & \ldots & r_{\varphi(a)} \\ r_1 + a & \ldots & r_{\varphi(a)} + a \\ \vdots & \vdots & \vdots \\ r_1 + (b-1)a & \ldots & r_{\varphi(a)} + (b-1)a \end{array}.$$

To find the value of $\varphi(ab)$ we have to count those numbers in the table which are coprime to b, as well. Observe that each column of the above table forms a complete residue system modulo b. Indeed, on the one hand each column of the table contains b elements, and on the other hand, if say the elements in row j_1 and row j_2 of column i are in the same residue class modulo b, that is

$$r_i + j_1 a \equiv r_i + j_2 a \pmod{b}$$

holds, then using

$$j_1 a \equiv j_2 a \pmod{b}$$

and $\gcd(a, b) = 1$ we get

$$j_1 \equiv j_2 \pmod{b}.$$

However, this by $0 \leq j_1, j_2 \leq b-1$ gives $j_1 = j_2$. That is, each column of the table forms a complete residue system modulo b indeed. Thus every column contains $\varphi(b)$ numbers coprime to b. As the number of columns is just $\varphi(a)$, we get

$$\varphi(ab) = \varphi(a)\varphi(b),$$

and our theorem is proved. □

Remark 1.2.2 One can easily see that none of the functions d, σ, φ, μ is totally multiplicative.

The forthcoming theorems provide so-called explicit forms of the above functions. By this we mean formulas which directly give the values of the functions in question taken on an $n \in \mathbb{N}$, once we have the prime factorization of n.

We start our studies with the functions $\chi(n)$ and $\mu(n)$. In fact the 'explicit forms' of these functions immediately follow by their definitions.

Theorem 1.2.4 *Beside $\chi(1) = 0$, if $n \in \mathbb{N}$, $n > 1$ and the prime factorization of n is given by*

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r},$$

then

$$\chi(n) = r.$$

Proof. The statement is a simple consequence of the definition of $\chi(n)$. □

Theorem 1.2.5 *Beside $\nu(1) = 0$, if $n \in \mathbb{N}$, $n > 1$ and the prime factorization of n is given by*

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r},$$

then

$$\nu(n) = \alpha_1 + \dots + \alpha_r.$$

Proof. The theorem immediately follows from the definition of $\nu(n)$. □

Now we turn to the investigation of the functions d, σ, φ, μ. Since the value of $\mu(n)$ can be obtained from its definition directly, in the following we discuss only the functions d, σ, φ.

Theorem 1.2.6 *Beside $d(1) = 0$, if $n \in \mathbb{N}$, $n > 1$ and the prime factorization of n is given by*

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r},$$

then

$$d(n) = (\alpha_1 + 1) \dots (\alpha_r + 1).$$

Proof. We give two different proofs to the statement. The first one is based upon the definition and the multiplicativity of d, while the second one relies on a simple combinatorial observation.

First proof. Obviously $d(1) = 1$. If $n > 1$, then by the multiplicativity of d, by Theorem 1.1.4 we may assume that n is of the form $n = p^\alpha$. Then the divisors of n are just

$$1, p, \dots, p^\alpha,$$

whence

$$d(n) = \alpha + 1.$$

From this our theorem directly follows, as in case of

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

by the multiplicativity of d we have

$$d(n) = (\alpha_1 + 1) \dots (\alpha_r + 1).$$

Second proof. Clearly, $d(1) = 1$. Let $n > 1$, and let

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

be the prime factorization of n. Then the divisors of n are of the form

$$m = p_1^{\beta_1} \dots p_r^{\beta_r}$$

with

$$0 \le \beta_i \le \alpha_i \quad (i = 1, \dots, r).$$

So the number of the divisors of n is

$$(\alpha_1 + 1) \dots (\alpha_r + 1),$$

which proves our statement again. □

Theorem 1.2.7 *Beside $\sigma(1) = 0$, if $n \in \mathbb{N}$, $n > 1$ and the prime factorization of n is given by*

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r},$$

then

$$\sigma(n) = \frac{p_1^{\alpha_1+1} - 1}{p_1 - 1} \dots \frac{p_r^{\alpha_r+1} - 1}{p_r - 1}.$$

Proof. We can readily check that $\sigma(1) = 1$. Assuming $n > 1$, using the multiplicativity of σ and Theorem 1.1.4, we may assume that n is of the shape $n = p^\alpha$. Then the divisors of n are

$$1, p, \dots, p^\alpha.$$

Since these numbers form a geometric progression of $\alpha+1$ terms, with initial term 1 and quotient p, their sum is given by

$$\sigma(n) = 1 + p + \dots + p^\alpha = \frac{p^{\alpha+1} - 1}{p - 1}.$$

This implies our statement, since for

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

we have

$$\sigma(n) = \frac{p_1^{\alpha_1+1} - 1}{p_1 - 1} \dots \frac{p_r^{\alpha_r+1} - 1}{p_r - 1},$$

which was to be proved. □

Theorem 1.2.8 *Besides $\varphi(1) = 1$, if $n \in \mathbb{N}$, $n > 1$ and*

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

is the prime factorization of n, then

$$\varphi(n) = (p_1 - 1) \dots (p_r - 1) p_1^{\alpha_1 - 1} \dots p_r^{\alpha_r - 1}$$

holds.

Proof. Certainly, $\varphi(1) = 1$. If $n > 1$, then by Theorem 1.1.4 as φ is multiplicative, we may assume that n is of the form $n = p^\alpha$. Then among the numbers not greater than n, the ones *not* being coprime to n are precisely these:

$$p, 2p, \ldots, p^{\alpha-1}p.$$

Hence $\varphi(n) = p^\alpha - p^{\alpha-1}$. This already implies our statement, since by the multiplicativity of φ, if n is of the shape

$$n = p_1^{\alpha_1} \ldots p_r^{\alpha_r}$$

then we obtain

$$\varphi(n) = (p_1 - 1) \ldots (p_r - 1) p_1^{\alpha_1 - 1} \ldots p_r^{\alpha_r - 1}.$$

□

Remark 1.2.3 Let

$$n = p_1^{\alpha_1} \ldots p_r^{\alpha_r}$$

be the prime factorization of $n > 1$. Observe that then by

$$(p_i - 1) p_i^{\alpha_i - 1} = p_i^{\alpha_i} \left(1 - \frac{1}{p_i}\right) \quad (i = 1, \ldots, r)$$

the above theorem gives

$$\varphi(n) = n \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

1.3 Summation and Möbius-transform of an arithmetic function

In this section we deal with the summation and Möbius-transform of an arithmetic function.

Definition 1.3.1 *Let f be an arithmetic function. Then the function*

$$F(n) = \sum_{d|n} f(d)$$

is called the summation or the summatory function of f, and the function

$$G(n) = \sum_{d|n} \mu(d) f\left(\frac{n}{d}\right)$$

is the Möbius-transform of f.

Remark 1.3.1 If $d \mid n$, then the numbers d and n/d are called the complementary divisors of n. Obviously, in the above summations their roles can be interchanged, that is (with the previous notation)

$$F(n) = \sum_{d|n} f\left(\frac{n}{d}\right)$$

and

$$G(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) f(d)$$

hold.

The next theorem implies that summation and Möbius-transform are in fact the inverse transformations of each other on the set of arithmetic functions. To prove this theorem we shall need the following lemma, which shows that the summatory function of the Möbius function is the earlier defined function δ.

Lemma 1.3.1 *Let $n \in \mathbb{N}$. Then we have*

$$\sum_{d|n} \mu(d) = \delta(n) = \begin{cases} 1, & \text{if } n=1, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Obviously,

$$\sum_{d|1} \mu(d) = \mu(1) = 1.$$

Let now $n > 1$, with prime factorization

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}.$$

Then by the definition of the μ function we have

$$\sum_{d|n}\mu(d)=\sum_{\{i_1,\ldots,i_l\}\subseteq\{1,\ldots,r\}}\mu(p_{i_1}\ldots p_{i_l})=$$
$$=\binom{r}{0}-\binom{r}{1}+\ldots+(-1)^r\binom{r}{r}=\sum_{j=1}^{r}(-1)^j\binom{r}{j}=(1-1)^r=0,$$

which was to be proved. □

Theorem 1.3.1 *The Möbius-transform of the summation of an arithmetic function is the original function. Also, the summation of the Möbius transform of an arithmetic function is the original function again.*

Proof. Let f be an arithmetic function, and write F for the summatory function of f. Then for the Möbius-transform of F we have

$$\sum_{d|n}\mu(d)F\left(\frac{n}{d}\right)=\sum_{d|n}\mu(d)\sum_{c|\frac{n}{d}}f(c)=$$
$$=\sum_{cd|n}\mu(d)f(c)=\sum_{c|n}f(c)\sum_{d|\frac{n}{c}}\mu(d)=f(n).$$

Let now G be the Möbius-transform of f. Then for the summatory function of G, using the previous lemma we obtain

$$\sum_{d|n}G(d)=\sum_{d|n}G\left(\frac{n}{d}\right)=\sum_{d|n}\sum_{c|\frac{n}{d}}\mu\left(\frac{n}{cd}\right)f(c)=$$
$$=\sum_{cd|n}\mu\left(\frac{n}{cd}\right)f(c)=\sum_{c|n}f(c)\sum_{d|\frac{n}{c}}\mu\left(\frac{n}{cd}\right)=f(n).$$

This proves our theorem. □

Our next theorem is of great importance: it shows that the set of multiplicative arithmetic functions is closed with respect to the above two transformations.

Theorem 1.3.2 *The summatory function and the Möbius-transform of a multiplicative arithmetic function are also multiplicative.*

Proof. Let f be a multiplicative arithmetic function with summatory function F. Further, let $a, b \in \mathbb{N}$, $\gcd(a,b) = 1$. Then, using that f is multiplicative, we obtain

$$F(ab) = \sum_{d|ab} f(d) = \sum_{d_1|a} \sum_{d_2|b} f(d_1 d_2) = \sum_{d_1|a} f(d_1) \sum_{d_2|b} f(d_2) = F(a)F(b).$$

Thus F is also multiplicative. Let now G be the Möbius-transform of f. Since f is multiplicative, for any $a, b \in \mathbb{N}$, $\gcd(a,b) = 1$

$$G(ab) = \sum_{d|ab} \mu\left(\frac{ab}{d}\right) f(d) = \sum_{d_1|a} \sum_{d_2|b} \mu\left(\frac{ab}{d_1 d_2}\right) f(d_1 d_2) =$$

$$= \left(\sum_{d_1|a} \mu\left(\frac{a}{d_1}\right) f(d_1)\right) \left(\sum_{d_2|b} \mu\left(\frac{b}{d_2}\right) f(d_2)\right) = G(a)G(b)$$

holds. Thus G is also multiplicative. □

1.4 Further assertions concerning the function σ

We mention two things concerning the function σ. The first one is about the so-called perfect numbers.

Definition 1.4.1 *The solutions of the equation $\sigma(n) = 2n$ are called perfect numbers.*

Example 1.4.1 The numbers 6 and 28 are perfect numbers. Indeed,

$$1 + 2 + 3 + 6 = 12$$

and

$$1 + 2 + 4 + 7 + 14 + 28 = 56$$

hold.

Remark 1.4.1 Observe that the above identities can also be written as

$$1 + 2 + 3 = 6$$

and

$$1 + 2 + 4 + 7 + 14 = 28.$$

In fact, these representations explain the interest - starting from the ancient Greeks - about such numbers: these numbers are the ones which can be obtained as the sums of their divisors smaller than them.

There are two standard conjectures (widely believed, but not justified statements) concerning perfect numbers:

Conjecture 1. There are infinitely many perfect numbers.

Conjecture 2. There are no odd perfect numbers.

The other thing to mention is the k-th power divisor summatory function.

Definition 1.4.2 *Let k be a non-negative integer. Then*

$$\sigma_k(n) = \sum_{d|n} d^k \quad (n \in \mathbb{N})$$

is the k-th power divisor summatory function.

Remark 1.4.2 We clearly have $\sigma_0(n) = d(n)$ and $\sigma_1(n) = \sigma(n)$. We further mention that $\sigma_k(n)$ is multiplicative for any $k \geq 0$.

1.5 Further assertions concerning the function φ

About the function φ, first we recall the Euler-Fermat theorem (without its proof):

Theorem 1.5.1 (The Euler-Fermat theorem) *Let $m \in \mathbb{N}$, $m \geq 2$. Then for all $a \in \mathbb{Z}$, $\gcd(a, z) = 1$ we have*

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

As it is well-known, as a simple consequence of the Euler-Fermat theorem we obtain the so-called little Fermat theorem:

Corollary 1.5.1 (The little Fermat theorem) *Let p be a prime. Then for all $a \in \mathbb{Z}$, $p \nmid a$ we have*

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

Later on, we shall need the summatory function of the function φ. This is given by the following theorem.

Theorem 1.5.2 *For every $n \in \mathbb{N}$*

$$\sum_{d|n} \varphi(d) = n$$

holds, that is, the summatory functions of Euler's function φ is the function $F(n) = n$.

Proof. Let

$$F(n) = \sum_{d|n} \varphi(d).$$

As φ is a multiplicative arithmetic function, so by Theorem 1.3.2 the function F is also multiplicative. One can readily check that $F(1) = 1$. Thus by Theorem 1.1.4 it is sufficient to check the statement on prime power values. Let p be a prime, $\alpha \in \mathbb{N}$. Then we have

$$\begin{aligned} F(p^\alpha) = \varphi(1) + \varphi(p) + \varphi(p^2) + \ldots + \varphi(p^\alpha) = \\ = 1 + (p-1) + (p^2 - p) + \ldots + (p^\alpha - p^{\alpha-1}) = p^\alpha. \end{aligned}$$

This proves our statement, since if $n > 1$ with prime factorization

$$p_1^{\alpha_1} \cdots p_r^{\alpha_r},$$

then by the multiplicativity of F we have

$$F(n) = F(p_1^{\alpha_1} \ldots p_r^{\alpha_r}) = p_1^{\alpha_1} \ldots p_r^{\alpha_r} = n.$$

□

1.6 Average of an arithmetic function

In this section we investigate the averages of arithmetic functions. Our motivation is that while the values of the studied arithmetic functions taken on neighboring integers can differ a lot, we hope that taking the average will 'smooth' this irregular behavior. As we shall see, this is what happens indeed.

For our discussion we shall need some notions. First we introduce the concept of the average of an arithmetic function (in the obvious way).

Definition 1.6.1 *The average function $H(n)$ of an arithmetic function $h(n)$ is defined by*

$$H(n) = \frac{1}{n}\sum_{i=1}^{n} h(i) \quad (n \in \mathbb{N}).$$

For the description of the asymptotic behavior of an arithmetic function we shall need the following notion.

Definition 1.6.2 *Let f and g be real valued arithmetic functions. If the quotient f/g is convergent and tends to 1, that is*

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1,$$

then we say that the functions $f(n)$ and $g(n)$ are asymptotically equal. Notation: $f(n) \sim g(n)$.

Now we are ready to formulate statements concerning the behavior of the average functions of some well-known arithmetic functions. From these, we shall prove only the theorem concerning the $d(n)$ function. For this we shall need the following lemma, which establishes an important connection of the logarithm function and the partial sums of the harmonic series. Note that in number theory instead of ln we traditionally use log for the notation of the natural logarithm. In what follows, just as in the whole lecture notes, we also follow this convention.

Lemma 1.6.1 *For any $n \geq 1$*

$$\log(n) \leq \sum_{j=1}^{n} \frac{1}{j} \leq \log(n) + 1$$

holds.

Proof. Observe that for $1 \leq j \leq n-1$ we have

$$\frac{1}{j+1} \leq \frac{1}{x} \leq \frac{1}{j} \quad (j \leq x \leq j+1).$$

Thus

$$\sum_{j=1}^{n-1} \frac{1}{j+1} \leq \int_1^n \frac{1}{x}\,\mathrm{dx} \leq \sum_{j=1}^{n-1} \frac{1}{j},$$

whence we get

$$\sum_{j=1}^{n} \frac{1}{j} = 1 + \sum_{j=1}^{n-1} \frac{1}{j+1} \leq \int_1^n \frac{1}{x}\,\mathrm{dx} + 1.$$

As $\log x$ is a primitive function of $1/x$, the Newton-Leibniz formula gives

$$\int_1^n \frac{1}{x}\,\mathrm{dx} = \log n - \log 1 = \log n.$$

Hence

$$\sum_{j=1}^{n} \frac{1}{j} \leq \log(n) + 1$$

and also

$$\log(n) \leq \sum_{j=1}^{n} \frac{1}{j}$$

holds, which proves our assertion. □

Now we formulate and prove the theorem concerning the average function of the $d(n)$ function.

Theorem 1.6.1 *Let*

$$D(n) = \frac{1}{n} \sum_{i=1}^{n} d(i) \quad (n \in \mathbb{N})$$

be the average function of $d(n)$*. Then*

$$D(n) \sim \log n.$$

	1	2	3	4	5	6
1	1	0	0	0	0	0
2	1	1	0	0	0	0
3	1	0	1	0	0	0
4	1	1	0	1	0	0
5	1	0	0	0	1	0
6	1	1	1	0	0	1

Table 1.1: The values δ_{ij} for $n = 6$.

Proof. Let $n \in \mathbb{N}$ be arbitrary but fixed. Make an $n \times n$ table, with δ_{ij} in its i-th row and j-th column $(1 \le i, j \le n)$, where

$$\delta_{ij} = \begin{cases} 1, & \text{if } j \mid i, \\ 0, & \text{otherwise.} \end{cases}$$

For example, if $n = 6$ then our table is given by Table 1.1.

Observe that then $d(i)$ is just the number of ones in the i-th row of the table, that is

$$d(i) = \delta_{i1} + \cdots + \delta_{in} \quad (i = 1, \ldots, n),$$

and thus

$$nD(n) = \sum_{i=1}^{n} d(i) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \delta_{ij} \right). \tag{1.3}$$

The proof of our theorem relies on that this double sum will be evaluated by interchanging the order of summations. In other words, we count the number of ones in the table also column-wise. For this observe that the number of ones in the j-th column of the table, that is

$$\delta_{1j} + \cdots + \delta_{nj} \quad (j = 1, \ldots, n)$$

is just the number of multiples of j from the set $\{1, \ldots, n\}$. Hence

$$\left\lfloor \frac{n}{j} \right\rfloor \le \delta_{1j} + \cdots + \delta_{nj} \le \left\lceil \frac{n}{j} \right\rceil \quad (j = 1, \ldots, n),$$

where $\lfloor x \rfloor$ is the lower integral part (floor) of the real number x, and $\lceil x \rceil$ is the upper integral part (ceiling) of the real number x. Thus

$$\frac{n}{j} - 1 \leq \sum_{i=1}^{n} \delta_{ij} \leq \frac{n}{j} + 1 \quad (j = 1, \ldots, n),$$

whence

$$\sum_{j=1}^{n} \left(\frac{n}{j} - 1 \right) \leq \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \delta_{ij} \right) \leq \sum_{j=1}^{n} \left(\frac{n}{j} + 1 \right)$$

follows. However, then (1.3) implies that

$$\sum_{i=1}^{n} \frac{1}{j} - 1 \leq D(n) \leq \sum_{i=1}^{n} \frac{1}{j} + 1.$$

Thus Lemma 1.6.1 gives

$$\log(n) - 1 \leq D(n) \leq \log(n) + 2.$$

So

$$\lim_{n \to \infty} \frac{D(n)}{\log n} = 1$$

implying

$$D(n) \sim \log n,$$

which was to be proved. □

Remark 1.6.1 We give the asymptotic behavior of the average functions of certain other well-known arithmetic functions without proofs:

$$\frac{1}{n} \sum_{i=1}^{n} \chi(i) \sim \log \log n \sim \frac{1}{n} \sum_{i=1}^{n} \nu(i),$$

$$\frac{1}{n} \sum_{i=1}^{n} \sigma(i) \sim \frac{\pi^2}{6} n, \quad \frac{1}{n} \sum_{i=1}^{n} \varphi(i) \sim \frac{3}{\pi^2} n.$$

1.7 Exercises

Exercise 1.7.1 Select the additive and the multiplicative arithmetic functions from the list below, and decide whether they are totally additive or totally multiplicative:

a) $f(n) = \sqrt{n}$

b) $f(n) = -2 \lg n$

c) $f(n) = e^n$

d) $f(n) = (-1)^{n+1}$

e) $f(n) = (-1)^{n+2}$

f) $f(n) = \frac{(-1)^{n+1}+1}{2}$

g) $f(n) = n^2 + 1$

h) $f(n) = 10^{n^2+1}$

i) $f(n) = \alpha$, where α is the exponent of 2 in the prime factorization of n

j) $f(n) = 3^\beta$, where β is the exponent of 3 in the prime factorization of n

Exercise 1.7.2 Does there exist a multiplicative arithmetic function $f(n)$, which apart from $f(1) = 1$ assumes only negative values?

Exercise 1.7.3 Does there exist an additive arithmetic function $f(n)$, which apart from $f(1) = 0$ assumes only odd integral values?

Exercise 1.7.4 Let $f(n)$ and $g(n)$ be not identically zero multiplicative arithmetic functions. Prove that then $h(n) = f(n) + g(n)$ is not a multiplicative arithmetic function!

Exercise 1.7.5 Give the values of the summatory functions and Möbius-transforms of the functions

$$f(n) = n^2 - 1, \quad g(n) = 2^n, \quad h(n) = (-1)^{n+1}$$

at

$$n = 1,\ 4,\ 6,\ 12.$$

Exercise 1.7.6 Give an example for an arithmetic function, whose summatory function is additive!

Exercise 1.7.7 Give an example for an arithmetic function, whose Möbius-transform is not multiplicative!

Exercise 1.7.8 Decide whether the following functions have multiplicative summatory functions and Möbius-transforms:

a) $f(n) = n + 1$

b) $f(n) = (-1)^{n+1}$

c) $f(n) = n^2$

Exercise 1.7.9 Give the values of the functions $\chi(n)$, $\nu(n)$, $d(n)$, $\sigma(n)$, $\varphi(n)$, $\mu(n)$ at

$$n = 1,\ 6,\ 30,\ 100,\ 210.$$

Exercise 1.7.10 Give the values of the summatory functions and Möbius transforms of $\chi(n)$, $\nu(n)$, $d(n)$, $\sigma(n)$, $\varphi(n)$, $\mu(n)$ at

$$n = 1,\ 4,\ 6,\ 10.$$

Exercise 1.7.11 Find the smallest positive integer n for which the value of $d(n)$ is

a) 23,

b) 25,

c) 24.

Exercise 1.7.12 Describe those positive integers n, which have the same number of odd and even divisors!

Exercise 1.7.13 Describe those positive integers n, for which $d(n)$ is odd!

Exercise 1.7.14 Prove that the equation

$$2d(n^2) = 3d(n)$$

holds if and only if n is a prime!

Exercise 1.7.15 Prove that

$$d(n+1) \geq 2d(n)$$

holds for infinitely many positive integer n.

Exercise 1.7.16 Show that for any $a, b \in \mathbb{N}$

$$d(ab) \leq d(a)d(b)$$

holds, and here we have equality if and only if $\gcd(a, b) = 1$.

Exercise 1.7.17 Prove that for any $a \in \mathbb{N}$ the set

$$A = \{n \in \mathbb{N} : a \mid d(n)\}$$

contains arbitrary long arithmetic progressions!

Exercise 1.7.18 Find those numbers $n \in \mathbb{N}$ for which $\sigma(n)$ is odd!

Exercise 1.7.19 Solve the equation

$$\sigma(n) = n + 3.$$

Exercise 1.7.20 Prove that for any $n \in \mathbb{N}$

$$\sigma(n) \leq \binom{n+1}{2}$$

holds!

Exercise 1.7.21 Prove that for any $a, b \in \mathbb{N}$

$$\sigma(ab) \leq \sigma(a)\sigma(b)$$

holds, and here we have equality if and only if $\gcd(a, b) = 1$.

Exercise 1.7.22 Can a prime be a perfect number?

Exercise 1.7.23 Prove that if

$$\sum_{d|n} \frac{1}{d} = 2,$$

then n is a perfect number!

Exercise 1.7.24 Solve the equation $\varphi(n) = 1$.

Exercise 1.7.25 Find those values of n for which $\varphi(n)$ is odd!

Exercise 1.7.26 Prove that for infinitely many even values of k, the equation $\varphi(n) = k$ is not solvable!

Exercise 1.7.27 Solve the equation $\varphi(n) = 1210$.

Exercise 1.7.28 Solve the equation $\varphi(n) = n - 2$.

Exercise 1.7.29 Prove that for any $n \in \mathbb{N}$ we have

$$\varphi(n^2) = n\varphi(n).$$

Exercise 1.7.30 Prove that for any $a, b \in \mathbb{N}$

$$\varphi(ab) \geq \varphi(a)\varphi(b)$$

holds, and here we have equality if and only if $\gcd(a, b) = 1$.

Exercise 1.7.31 Give the value of $\mu(n!)$ $(n \geq 0)$.

Exercise 1.7.32 Solve the following equations:

a) $\mu^2(n) + 2 = \mu(n)$,

b) $\mu^3(n) - \mu(n) = 0$,

c) $\mu(n^2) + \mu(n) = 2$,

d) $\mu(n) + 2 = \mu(6n)$.

Exercise 1.7.33 Prove that for every $n \in \mathbb{N}$

$$\mu(n^3) = \mu(n^2)$$

holds!

Exercise 1.7.34 Show that for any odd $n \in \mathbb{N}$ we have

$$\mu(n^2 + 3) = 0.$$

Exercise 1.7.35 Prove that for any $k \in \mathbb{N}$ one can find an $n \in \mathbb{N}$, such that

$$\mu(n+1) = \ldots = \mu(n+k) = 0$$

holds!

Exercise 1.7.36 Prove that for any $n \in \mathbb{N}$ we have

$$d(n) + \varphi(n) \leq n + 1.$$

Exercise 1.7.37 Show that for any $n \in \mathbb{N}$

$$\varphi(n)d(n) \geq n$$

holds!

Exercise 1.7.38 Prove that for any $n \in \mathbb{N}$

$$\varphi(n)\sigma(n) \leq n^2$$

is valid!

Exercise 1.7.39 Show that for every $n \in \mathbb{N}$ we have

$$\mu(n)(\nu(n) - \chi(n)) = 0.$$

Exercise 1.7.40 Show that for any $n \in \mathbb{N}$

$$2^{\chi(n)} \leq d(n) \leq 2^{\nu(n)}$$

holds!

Exercise 1.7.41 Prove that for any positive number K there exists an $n \in \mathbb{N}$, for which

$$\chi(n-1) - \chi(n) > K \quad \text{and} \quad \chi(n+1) - \chi(n) > K$$

hold!

Exercise 1.7.42 Prove that for infinitely many $n \in \mathbb{N}$ we have

$$\chi(n) > \chi(n-1) \quad \text{and} \quad \chi(n) > \chi(n+1).$$

Exercise 1.7.43 Show that for every positive number K one can find an $n \in \mathbb{N}$, such that

$$\nu(n-1) - \nu(n) > K \quad \text{and} \quad \nu(n+1) - \nu(n) > K$$

hold!

Exercise 1.7.44 Show that there exist infinitely many $n \in \mathbb{N}$ with

$$\nu(n) > \nu(n-1) \quad \text{and} \quad \nu(n) > \nu(n+1).$$

Exercise 1.7.45 Prove that for any positive number K there exists an $n \in \mathbb{N}$, for which

$$d(n) - d(n-1) > K \quad \text{and} \quad d(n) - d(n+1) > K$$

hold!

Exercise 1.7.46 Show that for every positive number K one can find an $n \in \mathbb{N}$, such that

$$\sigma(n-1) - \sigma(n) > K \quad \text{and} \quad \sigma(n+1) - \sigma(n) > K.$$

Exercise 1.7.47 Prove that for any positive number K there exists an $n \in \mathbb{N}$, for which

$$\varphi(n) - \varphi(n-1) > K \quad \text{and} \quad \varphi(n) - \varphi(n+1) > K$$

hold!

Exercise 1.7.48 Show that for infinitely many $n \in \mathbb{N}$ we have

$$\mu(n) < \min(\mu(n-1), \mu(n+1)).$$

Exercise 1.7.49 Prove that there are infinitely many $n \in \mathbb{N}$ with

$$\mu(n) > \max(\mu(n-1), \mu(n+1)).$$

Exercise 1.7.50 Show that for any $k \in \mathbb{N}$ there exists an $n \in \mathbb{N}$, for which

$$\mu(n-k) = \cdots = \mu(n-1) = \mu(n+1) = \cdots = \mu(n+k) = 0 \quad \text{and} \quad \mu(n) = 1$$

hold!

Exercise 1.7.51 Prove that for every $k \in \mathbb{N}$ one can find an $n \in \mathbb{N}$, such that

$$\mu(n-k) = \cdots = \mu(n-1) = \mu(n+1) = \cdots = \mu(n+k) = 0 \quad \text{and} \quad \mu(n) = -1.$$

Chapter 2

Order and generators modulo m

In this chapter we study the multiplicative structure of the modulo m residue class ring. We note that by a well-known theorem of Lagrange from basic algebra (or by its consequence that the order of any element of a finite group divides the order of the group), the theory developed in this chapter could be discussed in a more general framework.

2.1 The modulo m order and its basic properties

In this section we investigate the modulo m order and some of its fundamental properties.

Definition 2.1.1 *Let $m \geq 2$ be an integer, and let $a \in \mathbb{Z}$ be such that $\gcd(a, m) = 1$. Then by the order of a modulo m we mean the smallest positive integer d, for which*

$$a^d \equiv 1 \pmod{m}$$

holds. Notation: $\mathrm{ord}_m(a) = d$.

Remark 2.1.1 The Euler-Fermat theorem implies that $\mathrm{ord}_m(a)$ exists, and $\mathrm{ord}_m(a) \leq \varphi(m)$ is valid. In particular, if $m = p$ is a prime, then $\mathrm{ord}_m(a) \leq p - 1$.

In fact, much more is true than that. Namely, we have the following relation.

Theorem 2.1.1 *Using the above notation,* $\mathrm{ord}_m(a) \mid \varphi(m)$ *holds. In particular, if* $m = p$ *is a prime, then* $\mathrm{ord}_m(a) \mid p - 1$.

Proof. Let $d = \mathrm{ord}_m(a)$. Then the Euler-Fermat theorem implies that

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

Hence d is well-defined, and we have $1 \leq d \leq \varphi(m)$. Let $\varphi(m) = qd + r$, where q and r are non-negative integers and $0 \leq r < d$. As

$$a^d \equiv 1 \pmod{m},$$

by $\gcd(a, m) = 1$

$$a^r \equiv a^r \cdot \left(a^d\right)^q \equiv a^{qd+r} \equiv a^{\varphi(m)} \equiv 1 \pmod{m}$$

is also valid. Thus by the definition of d and the choice of r we obtain $r = 0$. This proves the statement. □

The following lemma reveals an assertion which is interesting in itself, but which also will prove to be rather useful later on.

Lemma 2.1.1 *Using the previous notation, put* $d = \mathrm{ord}_m(a)$. *Then for any* $k \geq 0$ *we have*

$$\mathrm{ord}_m(a^k) = \frac{d}{\gcd(d, k)}.$$

In particular, $\mathrm{ord}_m(a^k) = d$ *if and only if* $\gcd(d, k) = 1$.

Proof. Set $t = \mathrm{ord}_m(a^k)$. Then t is the smallest positive integer with

$$a^{tk} \equiv 1 \pmod{m}.$$

Hence the definition of d shows that t is the smallest positive integer with

$$tk \equiv 0 \pmod{d}.$$

However, the latter congruence is equivalent to the congruence

$$t \equiv 0 \pmod{\frac{d}{\gcd(d, k)}}.$$

Thus

$$t = \frac{d}{\gcd(d, k)},$$

which proves our claim. This trivially implies the special assertion, since by the above equality $t = d$ if and only if $\gcd(d, k) = 1$. □

2.2 Generators modulo m

In this section we study the question that for which moduli m is the multiplicative structure of the modulo m residue class ring cyclic. In this - as well as later on - the following notion will play an important role.

Definition 2.2.1 *Let m and a be integers with $m \geq 2$ and $\gcd(a, m) = 1$. If $\mathrm{ord}_m(a) = \varphi(m)$, then the element a is called a generator modulo m.*

Remark 2.2.1 The importance of generators (if they exist) comes from the fact that their powers generate the multiplicative group of invertible elements modulo m. In other words, using the standard notation: if there exists a generator modulo m, then the Abelian group $(\mathbb{Z}_m^*, \cdot)$ is cyclic.

We also mention that the 'generating' property in fact concerns modulo m residue classes, not integers. By this we mean that if a is a generator modulo m and

$$a \equiv b \pmod{m}$$

holds, then certainly b is a generator modulo m, too.

Now we give a precise description of moduli m for which there exist generators. This description will prove to be useful later on, as well.

Theorem 2.2.1 *Let $m \geq 2$. Then there exists a generator modulo m if and only if*

$$m = 2, 4, p^\alpha, 2p^\alpha,$$

where p is an odd prime and $\alpha \geq 1$.

Proof. We split the proof into several cases.

A) Assume first that m is a power of 2. Clearly, if $m = 2$ then $a = 1$, while if $m = 4$ then $a = 3$ is a generator modulo m, respectively. Let now $m = 2^\alpha$, where $\alpha \geq 3$. We show that there exists no generator modulo m in this case. For this it is obviously sufficient to check that for any odd $a \in \mathbb{Z}$ (and $\alpha \geq 3$) we have

$$a^{\varphi(2^\alpha)/2} \equiv 1 \pmod{2^\alpha}. \tag{2.1}$$

We prove the latter statement by induction on α. In case of $\alpha = 3$ by

$$1^2 \equiv 3^2 \equiv 5^2 \equiv 7^2 \equiv 1 \pmod{8}$$

the assertion holds. Assume now that (2.1) is valid for some odd $a \in \mathbb{Z}$ and $\alpha \geq 3$. This implies that

$$a^{\varphi(2^\alpha)/2} = 1 + t2^\alpha$$

holds, where $t \in \mathbb{Z}$. Taking the squares of both sides, this gives

$$a^{\varphi(2^\alpha)} = 1 + t2^{\alpha+1} + t^2 2^{2\alpha},$$

whence by $\alpha + 1 \leq 2\alpha$ we get

$$a^{\varphi(2^\alpha)} \equiv 1 \pmod{2^{\alpha+1}}.$$

Since

$$a^{\varphi(2^\alpha)} = a^{\varphi(2^{\alpha+1})/2},$$

we conclude that (2.1) is valid for the exponent $\alpha + 1$, as well. That is, the theorem is valid for moduli m which are powers of 2.

B) Let now $m = p$ be an odd prime. We show that then there exists a generator modulo m. In fact we shall prove more: we show that the number of generators is just $\varphi(p-1)$. Let d be an arbitrary divisor of $p-1$. Assume that a is an element of order d modulo p, that is $d = \mathrm{ord}_p(a)$; and hence certainly

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

Observe that then the numbers

$$a, a^2, \ldots, a^d \tag{2.2}$$

are all distinct (incongruent) modulo p. Indeed, if

$$a^u \equiv a^v \pmod{p}$$

would hold for some $1 \leq u < v \leq d$, then by $\gcd(a, p) = 1$

$$a^{v-u} \equiv 1 \pmod{p}$$

would also hold, which by $1 < v - u < d$ would contradict the choice of d. Thus the above elements in (2.2) are indeed different modulo p. Since

$$(a^i)^d \equiv 1 \pmod{p}$$

is valid for every $i \in \{1, \ldots, d\}$, thus the d roots of the polynomial $x^d - 1$ modulo p are just those given by (2.2). That is, all the elements of order d

modulo p are of the shape a^i $(1 \le i \le d)$. On the other hand, by Lemma 2.1.1 we know that among these powers a^i precisely those will be of order d modulo p, for which $\gcd(d, i) = 1$ holds. So altogether we have proved that the elements of order d modulo p are precisely the powers a^i with $\gcd(d, i) = 1$ $(1 \le i \le d)$. That is, if there exists an element of order d modulo p, then their number is $\varphi(d)$. In other words, writing $f_p(d)$ for the number of elements of order d modulo p, more precisely

$$f_p(d) = |\{b : 1 \le b \le p-1,\ \mathrm{ord}_p(b) = d\}|,$$

we have

$$f_p(d) \in \{0, \varphi(d)\} \quad (1 \le d \le p-1).$$

Observe that any $b \in \{1, \dots, p-1\}$ has an order modulo p, and this order $\mathrm{ord}_p(b)$ by Theorem 2.1.1 satisfies

$$\mathrm{ord}_p(b) \mid p-1.$$

Thus necessarily

$$\sum_{d \mid p-1} f_p(d) = p-1.$$

On the other hand, by Theorem 1.5.2 we have

$$\sum_{d \mid p-1} \varphi(d) = p-1.$$

This is possible only if

$$f_p(d) = \varphi(d) \quad (d \mid p-1),$$

that is, if for any d with $d \mid p-1$ the number of elements of order d is $\varphi(d)$. However, then in particular, the number of elements of order $p-1$ is just $\varphi(p-1)$. This shows that there exist generators modulo p, and their number (from the set $\{1, \dots, p-1\}$) is precisely $\varphi(p-1)$. Moreover, if a is a generator then a^i is a generator if and only if $\gcd(p-1, i) = 1$. Thus our theorem is proved also for $m = p$ prime, furthermore, we could even give the number of generators modulo p.

C) Consider now the case $m = p^\alpha$, where p is an odd prime and $\alpha \ge 2$. This case can be handled only in several steps. First we show that the exists a generator a modulo p, for which

$$a^{p-1} \not\equiv 1 \pmod{p^2} \tag{2.3}$$

holds. For this, let a be an arbitrary generator modulo p. If (2.3) holds, then we are done. Otherwise $b = a + p$ is also a generator modulo p, and just because (2.3) does not hold we get

$$b^{p-1} \equiv (a+p)^{p-1} \equiv a^{p-1} + (p-1)a^{p-2}p \equiv a^{p-1} - pa^{p-2} \equiv 1 - pa^{p-2} \pmod{p^2}.$$

However, then

$$b^{p-1} \not\equiv 1 \pmod{p^2},$$

since $p^2 \nmid pa^{p-2}$. That is, (2.3) indeed holds for some generator a modulo p.

Let now a be a generator modulo p, for which (2.3) holds. We show that then a is a generator modulo p^α for all $\alpha \geq 1$. This is certainly valid for $\alpha = 1$. Let $\alpha \geq 2$, and put $t = \mathrm{ord}_{p^\alpha}(a)$. We need to check that $t = \varphi(p^\alpha)$. As

$$a^t \equiv 1 \pmod{p^\alpha},$$

we have

$$a^t \equiv 1 \pmod{p}$$

as well, whence

$$t = q\varphi(p) = q(p-1)$$

holds with some integer q. Since $t \mid \varphi(p^\alpha)$ and $\varphi(p^\alpha) = (p-1)p^{\alpha-1}$, from this we obtain

$$q \mid p^{\alpha-1}.$$

Thus as p is a prime, we get $q = p^\beta$ with some $0 \leq \beta \leq \alpha - 1$, and

$$t = (p-1)p^\beta.$$

Assume now that here $\beta \leq \alpha - 2$. Then $\varphi(p^{\alpha-1}) = (p-1)p^{\alpha-2}$ yields

$$t \mid \varphi(p^{\alpha-1}).$$

But then

$$a^{\varphi(p^{\alpha-1})} \equiv 1 \pmod{p^\alpha} \tag{2.4}$$

is valid.

We show that if a is a generator modulo p and (2.3) holds, then (2.4) (with $\alpha \geq 2$) cannot be valid. We do this by induction on α. The case $\alpha = 2$ is automatic, since then (2.4) is just the negation of (2.3). Assume now that (2.4) is not valid for some $\alpha \geq 2$. By the Euler-Fermat theorem we have

$$a^{\varphi(p^{\alpha-1})} \equiv 1 \pmod{p^{\alpha-1}},$$

so we can write

$$a^{\varphi(p^{\alpha-1})} = 1 + kp^{\alpha-1}$$

with some integer k. Since (2.4) is not valid, here we have $p \nmid k$. Taking the p-th powers of both sides of the above equation, we obtain

$$a^{\varphi(p^{\alpha})} = (1 + kp^{\alpha-1})^p = 1 + kp^{\alpha} + k^2\frac{p(p-1)}{2}p^{2(\alpha-1)} + rp^{3(\alpha-1)}$$

with some $r \in \mathbb{Z}$. Observe that by $\alpha \geq 2$ we have

$$2\alpha - 1 \geq \alpha + 1 \quad \text{and} \quad 3\alpha - 3 \geq \alpha + 1.$$

Thus the above equality yields

$$a^{\varphi(p^{\alpha})} \equiv 1 + kp^{\alpha} \pmod{p^{\alpha+1}}.$$

As $p \nmid k$, we get

$$a^{\varphi(p^{\alpha})} \not\equiv 1 \pmod{p^{\alpha+1}}.$$

In other words, (2.4) is not valid also with $\alpha + 1$, and our claim follows.

Combining the above assertions, we conclude that $\beta \leq \alpha - 2$ is not possible. Thus necessarily

$$\beta = \alpha - 1,$$

whence

$$t = (p-1)p^{\alpha-1}.$$

This implies

$$\mathrm{ord}_{p^{\alpha}}(a) = t = \varphi(p^{\alpha}),$$

which proves our statement in the case $m = p^{\alpha}$, as well.

D) Suppose now that m is of the shape $m = 2p^{\alpha}$, where p is an odd prime and $\alpha \geq 1$. Let a be an odd generator modulo p^{α}. (We have such an element, as if a is an even generator modulo p^{α} then $a + p^{\alpha}$ is an appropriate choice.) Let $d = \mathrm{ord}_{2p^{\alpha}}(a)$. We show that $d = \varphi(2p^{\alpha})$. For this observe that

$$d \mid \varphi(2p^{\alpha})$$

and

$$\varphi(2p^{\alpha}) = \varphi(2)\varphi(p^{\alpha}) = \varphi(p^{\alpha})$$

imply

$$d \mid \varphi(p^{\alpha}).$$

On the other hand, by

$$a^d \equiv 1 \pmod{2p^\alpha}$$

we obtain

$$a^d \equiv 1 \pmod{p^\alpha},$$

whence

$$\varphi(p^\alpha) \mid d$$

follows. Thus

$$d = \varphi(p^\alpha) = \varphi(2p^\alpha),$$

that is, a is a generator modulo $2p^\alpha$. So the statement follows also for $m = 2p^\alpha$.

E) Finally, assume that m does not belong to any of the cases investigated so far. We prove that then there are no generators modulo m. Let

$$m = 2^\alpha p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

be the prime factorization of m, where $\alpha \geq 0$ and $\alpha_i > 0$ $(i = 1, \dots, r)$. Observe that by the choice of m we further have $\alpha \geq 2$ or $r \geq 2$. In particular, $\varphi(m)$ is even. We prove that for any $a \in \mathbb{Z}$, $\gcd(a, m) = 1$ we have

$$a^{\varphi(m)/2} \equiv 1 \pmod{m}.$$

This will certainly prove our statement, since then

$$\mathrm{ord}_m(a) < \varphi(m).$$

For this, let a be an arbitrary integer coprime to m, and let g be a generator modulo $p_1^{\alpha_1}$. Choose a k such that

$$a \equiv g^k \pmod{p_1^{\alpha_1}}.$$

Then

$$a^{\varphi(m)/2} \equiv g^{k\varphi(m)/2} \equiv g^{t\varphi(p_1^{\alpha_1})} \pmod{p_1^{\alpha_1}},$$

where

$$t = k\varphi(2^\alpha)\varphi(p_2^{\alpha_2}) \dots \varphi(p_r^{\alpha_r})/2.$$

Here t is an integer, since in case of $\alpha < 2$ we have $r \geq 2$. However, then using the above congruence, by the Euler-Fermat theorem we obtain

$$a^{\varphi(m)/2} \equiv 1 \pmod{p_1^{\alpha_1}}.$$

A similar argument yields that in fact for any $i = 1, \ldots, r$

$$a^{\varphi(m)/2} \equiv 1 \pmod{p_i^{\alpha_i}} \tag{2.5}$$

holds. We show that the above congruence is valid also modulo 2^α. If $\alpha \geq 3$, then by what we have proved in part A) we get that

$$a^{\varphi(2^\alpha)/2} \equiv 1 \pmod{2^\alpha}.$$

As $\varphi(2^\alpha) \mid \varphi(m)$, this gives

$$a^{\varphi(m)/2} \equiv 1 \pmod{2^\alpha} \tag{2.6}$$

indeed. If $\alpha \leq 2$, then by the Euler-Fermat we obtain

$$a^{\varphi(2^\alpha)} \equiv 1 \pmod{2^\alpha}.$$

As $r \geq 1$, we have

$$\varphi(m) = \varphi(2^\alpha)\varphi(p_1^{\alpha_1}) \ldots \varphi(p_r^{\alpha_r}) = 2s\varphi(2^\alpha)$$

with some integer s. Therefore (2.6) holds once again. Combining the congruences (2.5) and (2.6),

$$a^{\varphi(m)/2} \equiv 1 \pmod{m}$$

follows. Thus the proof of the theorem is complete. □

The last theorem of this section gives the number of generators.

Theorem 2.2.2 *Let $m \geq 2$. If there exists a generator modulo m, then the number of distinct (incongruent) generators modulo m is $\varphi(\varphi(m))$. Furthermore, if a is a generator modulo m, then a^n $(1 \leq n \leq \varphi(m))$ is a generator modulo m if and only if $\gcd(n, \varphi(m)) = 1$.*

Proof. Let a be a generator modulo m. Then

$$\mathrm{ord}_m(a) = \varphi(m).$$

Thus by Lemma 2.1.1 we obtain

$$\mathrm{ord}_m(a^n) = \frac{\varphi(m)}{\gcd(n, \varphi(m))}.$$

Hence the order of a^n is $\varphi(m)$, or in other words, a^n is a generator modulo m if and only if $\gcd(n, \varphi(m)) = 1$. As the number of such powers of a is just $\varphi(\varphi(m))$, and the numbers

$$a, a^2, \ldots, a^{\varphi(m)}$$

form a reduced residue system modulo m, our statement follows. □

2.3 The index calculus

In this section we discuss the index calculus, which is in fact a discrete version of logarithm.

Let m be such that there exists a generator modulo m, and let a be a generator modulo m. Then the numbers

$$1, a, \ldots, a^{\varphi(m)-1}$$

form a reduced residue system modulo m. This provides an opportunity to introduce the following notion.

Definition 2.3.1 *Let a be a generator modulo m and let $b \in \mathbb{Z}$, $\gcd(b, m) = 1$. Then there exists a uniquely determined integer k, for which*

$$a^k \equiv b \pmod{m} \quad \textit{and} \quad 1 \leq k \leq \varphi(m) - 1$$

hold. This exponent k is called the index of b with respect to a modulo m. Notation: $\mathrm{ind}_a(b) = k$.

Some fundamental properties of the index are established by the following theorem.

Theorem 2.3.1 *Let a be a generator modulo m, and let b_1, b_2 be integers with $\gcd(b_1, m) = \gcd(b_2, m) = 1$. Then the following assertions hold:*

i) $\mathrm{ind}_a(b_1 b_2) \equiv \mathrm{ind}_a(b_1) + \mathrm{ind}_a(b_2) \pmod{\varphi(m)}$,

ii) for any $n \geq 1$ we have $\mathrm{ind}_a(b_1^n) \equiv n \cdot \mathrm{ind}_a(b_1) \pmod{\varphi(m)}$,

iii) $\mathrm{ind}_a(1) = 0$ *and* $\mathrm{ind}_a(a) = 1$,

iv) if $m > 2$, then $\mathrm{ind}_a(-1) = \varphi(m)/2$,

v) if a' is a generator modulo m, then

$$\mathrm{ind}_a(b_1) \equiv \mathrm{ind}_{a'}(b_1) \cdot \mathrm{ind}_a(a') \pmod{\varphi(m)}.$$

Proof. We verify our statements separately.

i) Let $k_1 = \mathrm{ind}_a(b_1)$ and $k_2 = \mathrm{ind}_a(b_2)$. Then of course

$$a^{k_1} \equiv b_1 \pmod{m} \quad \text{and} \quad a^{k_2} \equiv b_2 \pmod{m},$$

whence

$$a^{k_1+k_2} \equiv b_1 b_2 \pmod{m}.$$

Thus

$$\mathrm{ind}_a(b_1 b_2) \equiv k_1 + k_2 \pmod{\varphi(m)},$$

which was to be proved.

ii) Let again $k_1 = \mathrm{ind}_a(b_1)$. Then

$$a^{k_1} \equiv b_1 \pmod{m},$$

and thus for any $n \geq 1$ we have

$$a^{nk_1} \equiv b_1^n \pmod{m}.$$

From this we get

$$\mathrm{ind}_a(b_1^n) \equiv nk_1 \pmod{\varphi(m)},$$

and our claim follows.

iii) In view of

$$a^0 \equiv 1 \pmod{m} \quad \text{and} \quad a^1 \equiv a \pmod{m}$$

the statement trivially holds.

iv) Let $k = \mathrm{ord}_a(-1)$. Then

$$a^k \equiv -1 \pmod{m},$$

whence we obtain

$$a^{2k} \equiv 1 \pmod{m}.$$

But then

$$2k \equiv 0 \pmod{\varphi(m)}.$$

Since $1 \leq k \leq \varphi(m) - 1$, this implies

$$k = \frac{\varphi(m)}{2}$$

and the statement follows.

v) Let $\mathrm{ind}_{a'}(b_1) = k_1'$ and $\mathrm{ind}_a(a') = k'$. Then of course

$$(a')^{k_1'} \equiv b_1 \pmod{m} \quad \text{and} \quad a^{k'} \equiv a' \pmod{m}.$$

So

$$b_1 \equiv (a')^{k'_1} \equiv (a^{k'})^{k'_1} \equiv a^{k'k'_1} \pmod{m}.$$

Hence

$$\mathrm{ind}_a(b_1) \equiv k'k'_1 \pmod{\varphi(m)},$$

which proves our claim. □

Remark 2.3.1 The above theorem shows that the index indeed holds properties similar to those of the logarithm. This is the reason why $\mathrm{ind}_a(b)$ is often referred to as the discrete logarithm of b with respect to a.

2.4 Higher order and exponential congruences

In this section, as a kind of application of index calculus, we discuss certain polynomial and exponential congruences.

We start with the case of linear congruences.

Theorem 2.4.1 *Let $m \geq 2$ be such that there exists a generator modulo m, and let a be a generator modulo m. Let u and v be integers such that $\gcd(u, m) = \gcd(v, m) = 1$. Consider the linear congruence*

$$ux \equiv v \pmod{m}, \tag{2.7}$$

where x is an unknown integer. Then the unique solution modulo m of this congruence is given by

$$x \equiv a^{\mathrm{ind}_a(v) - \mathrm{ind}_a(u)} \pmod{m}.$$

Proof. It is well-known that by $\gcd(u, m) = 1$ the linear congruence in the statement is solvable, and has only one solution modulo m. Observe that taking the indices of both sides of (2.7) with respect to a, by the properties of the index (2.7) is in fact equivalent with the congruence

$$\mathrm{ind}_a(u) + \mathrm{ind}_a(x) \equiv \mathrm{ind}_a(v) \pmod{\varphi(m)}.$$

After rearrangement, the above congruence yields

$$\mathrm{ind}_a(x) \equiv \mathrm{ind}_a(v) - \mathrm{ind}_a(u) \pmod{\varphi(m)}.$$

From this - raising a on both sides above - our statement immediately follows. $\square$

Example 2.4.1 Let $m = 7$, $a = 3$. Consider the congruence

$$2x \equiv 5 \pmod{7}$$

Now

$$\mathrm{ind}_3(2) = 2 \quad \text{and} \quad \mathrm{ind}_3(5) = 5.$$

Thus by

$$\mathrm{ind}_3(x) = \mathrm{ind}_3(5) - \mathrm{ind}_3(2) = 3$$

we get

$$x \equiv 3^3 \equiv 6 \pmod{7}.$$

Our next theorem concerns a specific type of polynomial congruences, so-called binom congruences.

Theorem 2.4.2 *Let $m \geq 2$ be such that there exists a generator modulo m, and let a be a generator modulo m. Let $n \geq 1$, $b \in \mathbb{Z}$, $\gcd(b, m) = 1$, and consider the congruence*

$$x^n \equiv b \pmod{m}, \tag{2.8}$$

where x is an unknown integer. Then the congruence is solvable if and only if

$$\gcd(n, \varphi(m)) \mid \mathrm{ind}_a(b).$$

Further, if the above divisibility relation holds, then the number of solutions of (2.8) modulo m is given by

$$\gcd(n, \varphi(m)).$$

Proof. Observe that (2.8) - taking the index of both sides, and using the properties of the index - is equivalent to the congruence

$$n \cdot \mathrm{ind}_a(x) \equiv \mathrm{ind}_a(b) \pmod{\varphi(m)}.$$

This is a linear congruence for the unknown value $\mathrm{ind}_a(x)$, which is solvable if and only if $\gcd(n, \varphi(m)) \mid \mathrm{ind}_a(b)$. This already proves the first part of our statement. Further, we also know that if the divisibility relation holds, then

the number of solutions of the above linear congruence modulo $\varphi(m)$ is just $\gcd(n, \varphi(m))$. This proves the second part of the statement. $\square$

Example 2.4.2 Let $n = 13$, $a = 2$. Consider the congruence

$$x^3 \equiv -1 \pmod{13}.$$

Now by

$$\mathrm{ind}_2(-1) = 6$$

we obtain the congruence

$$3 \cdot \mathrm{ind}_2(x) \equiv 6 \pmod{12}.$$

As $\gcd(3, 12) = 3$ and $3 \mid 6$, this is solvable, and the number of solutions is $\gcd(3, 12) = 3$. Actually solving the congruence, we get

$$\mathrm{ind}_2(x) = 2, 6, 10,$$

whence

$$x \equiv 2^2, 2^6, 2^{10} \equiv 4, 12, 10 \pmod{13}$$

follows.

We conclude the section by discussing exponential congruences.

Theorem 2.4.3 *Let $m \geq 2$ be such that there exists a generator modulo m, and let a be a generator modulo m. Let u and v be integers with $\gcd(u, m) = \gcd(v, m) = 1$, and consider the congruence*

$$u^x \equiv v \pmod{m}, \tag{2.9}$$

where x is an unknown integer. Then the congruence is solvable if and only if

$$\gcd(\mathrm{ind}_a(u), \varphi(m)) \mid \mathrm{ind}_a(v).$$

Further, if the above divisibility relation holds, then the number of solutions of (2.9) modulo m is given by

$$\gcd(\mathrm{ind}_a(u), \varphi(m)).$$

Proof. Taking the index of both sides of the congruence (2.9), using the properties of the index we obtain that (2.9) is equivalent to the congruence

$$x \cdot \mathrm{ind}_a(u) \equiv \mathrm{ind}_a(v) \pmod{\varphi(m)}.$$

This is a linear congruence for the unknown integer x, which is solvable if and only if

$$\gcd(\mathrm{ind}_a(u), \varphi(m)) \mid \mathrm{ind}_a(v).$$

This proves the first part of the theorem. Further, we also know that if the above divisibility holds, then the number of solutions of our congruence modulo $\varphi(m)$ is just $\gcd(\mathrm{ind}_a(u), \varphi(m))$. Thus the proof of the statement is complete. □

Example 2.4.3 Let $m = 11$, $a = 7$. Consider the exponential congruence

$$3^x \equiv 4 \pmod{11}.$$

Now

$$\mathrm{ind}_7(3) = 4 \quad \text{and} \quad \mathrm{ind}_7(4) = 6.$$

Thus taking the indices in the above congruence with respect to 7, we get the linear congruence

$$4x \equiv 6 \pmod{10}.$$

Since $\gcd(4, 10) = 2$ and $2 \mid 6$ this congruence is solvable, and has two solutions modulo 10. In particular, we obtain

$$x \equiv 4, 9 \pmod{10}.$$

2.5 Exercises

Exercise 2.5.1 Give the orders of the following numbers modulo 31:

a) 2,

b) 3,

c) 71.

Exercise 2.5.2 We know that

$$18^{15} \equiv 1 \pmod{41}$$

and

$$18^{25} \equiv 1 \pmod{41}.$$

What is the order of 18 modulo 41 ?

Exercise 2.5.3 Let p be a prime, and a and b be integers with $p \nmid ab$. Prove that if

$$a \equiv b^k \pmod{p}$$

and

$$b \equiv a^\ell \pmod{p}$$

hold with some positive integers k, ℓ, then

$$\mathrm{ord}_p(a) = \mathrm{ord}_p(b).$$

Exercise 2.5.4 Let p be a prime, $a \in \mathbb{Z}$, $p \nmid a$. Give the remainder of

$$\sum_{i=1}^{\mathrm{ord}_p(a)} a^i$$

after division by p.

Exercise 2.5.5 Let p be a prime, and a and b be integers with $p \nmid ab$. Prove that then

$$\mathrm{ord}_p(ab) \leq \mathrm{lcm}(\mathrm{ord}_p(a), \mathrm{ord}_p(b)).$$

Give an example where

$$\mathrm{ord}_p(ab) < \mathrm{lcm}(\mathrm{ord}_p(a), \mathrm{ord}_p(b))$$

holds.

Exercise 2.5.6 Determine whether the following numbers are generators modulo 23 or not:

a) 5,

b) 6,

c) 7.

Exercise 2.5.7 Find the smallest positive integer which is a generator modulo p, where

a) $p = 7$,

b) $p = 11$,

c) $p = 13$,

d) $p = 17$.

Exercise 2.5.8 Find the smallest positive integer which is a generator modulo m, where

a) $m = 14$,

b) $m = 22$,

c) $m = 26$,

d) $m = 34$,

e) $m = 27$,

f) $m = 125$,

g) $m = 54$,

h) $m = 98$.

Exercise 2.5.9 Find those primes p, for which $p - 1$ is a generator modulo p.

Exercise 2.5.10 Find those moduli m with $2 \leq m \leq 101$, for which the following number a is a generator modulo m:

a) $a = 2$,

b) $a = 3$,

c) $a = 4$,

d) $a = 5$,

e) $a = 7$,

f) $a = 11$,

g) $a = -1$,

h) $a = -2$.

Exercise 2.5.11 Let p be a prime and $a \in \mathbb{Z}$. Prove that if a^k is a generator modulo p for some positive integer k, then so is a.

Exercise 2.5.12 Let $p = 2q + 1$, where p and q are odd primes. Prove that then for any $a \in \mathbb{Z}$ the congruence

$$a^3 - a \not\equiv 0 \pmod{p}$$

implies that one of the numbers a and $-a$ is a generator modulo p.

Exercise 2.5.13 Let p be an odd prime and a a generator modulo p. Give the value of the index

$$\mathrm{ind}_a(p-1).$$

Exercise 2.5.14 We know that 2 is a generator modulo 11. Using this, solve the following linear congruences in unknown integers x:

- $2x \equiv 3 \pmod{11}$,
- $3x \equiv 7 \pmod{11}$,
- $7x \equiv 2 \pmod{11}$.

Exercise 2.5.15 Solve the following binom congruences in integers x:

- $x^3 \equiv 1 \pmod{13}$,

- $x^{14} \equiv 1 \pmod{23}$,
- $x^4 \equiv 2 \pmod{11}$,
- $x^{10} \equiv 5 \pmod{37}$,
- $4x^6 \equiv 13 \pmod{23}$,
- $19x^8 \equiv x^{16} \pmod{41}$.

Exercise 2.5.16 Solve the following exponential congruences in non-negative integers x:

- $6^x \equiv 7 \pmod{11}$,
- $5^x \equiv 7 \pmod{13}$,
- $4 \cdot 9^x \equiv 11 \pmod{19}$,
- $11 \cdot 9^{5x+1} \equiv 19 \pmod{23}$.

Chapter 3

Quadratic residues

In this section we study quadratic residues modulo p. Related to this, we are going to investigate the Legendre symbol and its properties, followed by the discussion of the Jacobi symbol and its properties.

3.1 Quadratic residues modulo p

First we study quadratic residues modulo p.

Definition 3.1.1 *Let $p > 2$ be a prime and let $a \in \mathbb{Z}$ such that $p \nmid a$. If we have*

$$x^2 \equiv a \pmod{p},$$

with some integer x, then we say that a is a quadratic residue, and the residue class $a \pmod{p}$ is called a quadratic residue class modulo p. Otherwise, we say that a is a quadratic non-residue, and $a \pmod{p}$ is a quadratic non-residue class modulo p.

Our first theorem describes the number of quadratic residue classes modulo p.

Theorem 3.1.1 *Let p be an odd prime. Then the number of quadratic residue classes modulo p is $(p-1)/2$. Further, if a is a generator modulo p, then a^i $(1 \leq i \leq p-1)$ is a quadratic residue modulo p if and only if i is even.*

Proof. Since

$$x^2 \equiv (-x)^2 \pmod{p},$$

a set of representatives of quadratic residue classes is given by

$$1^2, 2^2, \ldots, \left(\frac{p-1}{2}\right)^2.$$

Moreover, these residue classes are pairwise distinct, as for $1 \leq i < j \leq (p-1)/2$ the congruence

$$i^2 \equiv j^2 \pmod{p}$$

cannot hold. Indeed, otherwise $p \mid j^2 - i^2$, and hence $p \mid j + i$ or $p \mid j - i$ would be valid, but by

$$1 \leq j - i < j + i < p$$

it is impossible. This proves our first statement.

To justify our second statement, let $1 \leq i \leq p-1$ be even, $i = 2j$ say. Then of course

$$a^i \equiv (a^j)^2 \pmod{p},$$

so a^i is a quadratic residue modulo p indeed. On the other hand, if a^i is a quadratic residue modulo p for some $1 \leq i \leq p-1$, then

$$a^i \equiv (a^j)^2 \pmod{p}$$

holds with some $1 \leq j \leq p-1$. Thus

$$i \equiv 2j \pmod{p-1},$$

whence i is even. Hence the proof of the theorem is complete. □

3.2 The Legendre symbol

In this section we introduce the Legendre symbol and discuss its basic properties.

Definition 3.2.1 *Let p be a prime, $p > 2$ and a be an arbitrary integer. Then the Legendre symbol $\left(\frac{a}{p}\right)$ is defined in the following way:*

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & \text{if } p \mid a, \\ 1, & \text{if } a \text{ is a quadratic residue} \pmod{p}, \\ -1, & \text{if } a \text{ is a quadratic non-residue} \pmod{p}. \end{cases}$$

Our next theorem gives a very useful property of the Legendre symbol.

Theorem 3.2.1 *Let p be an odd prime, $a \in \mathbb{Z}$. Then*

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}.$$

Proof. If $p \mid a$, then the statement is trivial. Assume that $p \nmid a$. Then we have

$$\left(a^{\frac{p-1}{2}}\right)^2 \equiv 1 \pmod{p},$$

whence

$$a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}.$$

Let g be a generator modulo p, and let i be the exponent for which

$$a \equiv g^i \pmod{p} \quad (1 \leq i \leq p-1).$$

Then a is a quadratic residue if and only if i is even, that is if and only if

$$g^{\frac{i(p-1)}{2}} \equiv 1 \pmod{p},$$

since

$$g^t \equiv 1 \pmod{p}$$

is equivalent to $p-1 \mid t$. This proves our statement. □

Our next theorem describes some further properties of the Legendre symbol.

Theorem 3.2.2 *Let p be an odd prime. Then*

$$\left(\frac{1}{p}\right) = 1, \quad \left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}.$$

Further, for any $a, b \in \mathbb{Z}$ we have

a) if $a \equiv b \mod p$, then

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right),$$

b) the Legendre symbol is multiplicative, that is

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right),$$

c) if $\gcd(b,p) = 1$*, then*

$$\left(\frac{ab^2}{p}\right) = \left(\frac{a}{p}\right).$$

Proof. The first statement is trivial, and the second assertion is a consequence of the first one. Further, a) is an immediate consequence of the definition, while b) follows from the congruences

$$\left(\frac{ab}{p}\right) \equiv (ab)^{\frac{p-1}{2}} \equiv a^{\frac{p-1}{2}} b^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \pmod{p}.$$

Finally, c) easily follows from b) and the definition. □

Now we formulate a lemma due to Gauss, which will prove to be very useful later on.

Lemma 3.2.1 *Let p be an odd prime, $a \in \mathbb{Z}$, $p \nmid a$. Consider the smallest positive residues of the numbers*

$$a, 2a, \ldots \frac{p-1}{2}a \tag{3.1}$$

modulo p, and write n for the number of those in this list which are greater than $p/2$. Then

$$\left(\frac{a}{p}\right) = (-1)^n.$$

Proof. Let $u_1, \ldots, u_k$ be those smallest positive residues modulo p of the numbers (3.1) which are smaller than $p/2$, and $v_1, \ldots, v_n$ those which are greater than $p/2$. Hence clearly

$$k + n = \frac{p-1}{2},$$

and

$$u_1, \ldots, u_k, v_1, \ldots, v_n$$

are the smallest positive residues of the numbers ta $(1 \le t \le (p-1)/2)$. Obviously, the numbers

$$u_1, \ldots, u_k, p - v_1, \ldots, p - v_n \tag{3.2}$$

are all between 1 and $(p-1)/2$. We prove that these $(p-1)/2$ numbers are pairwise distinct. If

$$u_i = u_j$$

holds for some $1 \le i, j \le k$, then with some integers $1 \le t, s \le (p-1)/2$

$$ta \equiv sa \pmod{p}$$

holds. As $p \nmid a$, this implies $t = s$, thus $i = j$. In a similar way we obtain that

$$p - v_i = p - v_j \quad (1 \le i, j \le n)$$

yields $i = j$. Assume now that

$$u_i = p - v_j \quad (1 \le i \le k,\ 1 \le j \le n).$$

Let

$$u_i \equiv ta \pmod{p} \quad \left(1 \le t \le \frac{p-1}{2}\right)$$

and

$$v_j \equiv sa \pmod{p} \quad \left(1 \le s \le \frac{p-1}{2}\right).$$

Then

$$ta \equiv u_i \equiv p - v_j \equiv p - sa \equiv -sa \pmod{p}$$

implies

$$p \mid (t+s)a.$$

However, this by

$$p \nmid a \quad \text{and} \quad 1 \le t + s \le p - 1$$

is impossible. This shows that the numbers (3.2) are pairwise distinct indeed. Since these are $(p-1)/2$ positive integers, all of them at most $(p-1)/2$, we get that the numbers (3.2) are in fact

$$1, 2, \ldots, \frac{p-1}{2}$$

in some order. This yields that

$$\left(\frac{p-1}{2}\right)!a^{\frac{p-1}{2}} \equiv a \cdot 2a \cdot \ldots \cdot \frac{p-1}{2}a \equiv u_1 \cdot \ldots \cdot u_k \cdot v_1 \cdot \ldots \cdot v_n \equiv$$
$$\equiv (-1)^n u_1 \cdot \ldots \cdot u_k \cdot (p-v_1) \cdot \ldots \cdot (p-v_n) \equiv (-1)^n \left(\frac{p-1}{2}\right)! \pmod{p}.$$

From this, after canceling $((p-1)/2)!$ (which is not divisible by p)

$$a^{\frac{p-1}{2}} \equiv (-1)^n \pmod{p}$$

follows. As Theorem 3.2.1 shows that

$$a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p},$$

this proves our statement. □

Our next theorem discusses a special but important case.

Theorem 3.2.3 *Let p be a prime, $p > 2$. Then*

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$

Proof. We apply Lemma 3.2.1 with the choice $a = 2$. For this observe that among

$$2, 4, 6, \ldots, p-1$$

the number of those smaller than $p/2$ is $\lfloor (p-1)/4 \rfloor$, so the number of those among them which are greater than $p/2$ is

$$n = \frac{p-1}{2} - \left\lfloor \frac{p-1}{4} \right\rfloor.$$

(Here $\lfloor x \rfloor$ stands for the lower integer part of the real number x.) Thus we get

$$(-1)^n = \begin{cases} 1, & \text{if } p \equiv \pm 1 \pmod{8}, \\ -1, & \text{if } p \equiv \pm 3 \pmod{8}. \end{cases}$$

This by Lemma 3.2.1 proves our claim. □

The last theorem of the section because of its importance holds a particular name: this is the so-called quadratic reciprocity law.

Theorem 3.2.4 (Quadratic reciprocity) *Let p and q be odd primes. Then we have*

$$\left(\frac{q}{p}\right) = (-1)^{\frac{(p-1)(q-1)}{4}} \left(\frac{p}{q}\right).$$

Proof. We verify two assertions, which together imply our statement. First we show that if a is odd and $p \nmid a$, then

$$\left(\frac{a}{p}\right) = (-1)^{\ell} \tag{3.3}$$

holds, where

$$\ell = \sum_{t=1}^{\frac{p-1}{2}} \left\lfloor \frac{ta}{p} \right\rfloor.$$

For this we shall use Lemma 3.2.1, together with the information obtained and notation used in its proof. To verify the assertion (3.3), it is clearly sufficient to show that

$$\ell \equiv n \pmod 2.$$

Observe that a permutation of the numbers

$$a, 2a, \ldots, \frac{p-1}{2}a$$

is given by

$$\left\lfloor \frac{ta}{p} \right\rfloor p + u_1, \ldots, \left\lfloor \frac{ta}{p} \right\rfloor p + u_k, \left\lfloor \frac{ta}{p} \right\rfloor p + v_1, \ldots, \left\lfloor \frac{ta}{p} \right\rfloor p + v_n.$$

Thus after summation we get

$$\left(1 + 2 + \cdots + \frac{p-1}{2}\right) a = p \sum_{t=1}^{\frac{p-1}{2}} \left\lfloor \frac{ta}{p} \right\rfloor + \sum_{i=1}^{k} u_i + \sum_{j=1}^{n} v_j.$$

Using that the numbers

$$u_1, \ldots, u_k, p - v_1, \ldots, p - v_n$$

form a permutation of

$$1, 2, \ldots, \frac{p-1}{2},$$

after rearrangement hence we get

$$\left(1+2+\cdots+\frac{p-1}{2}\right)(a-1)+2\sum_{j=1}^{n}v_j=p\left(\sum_{t=1}^{\frac{p-1}{2}}\left\lfloor\frac{ta}{p}\right\rfloor+n\right)=p(\ell+n).$$

As a is odd, the above equality yields

$$\ell\equiv n \pmod 2,$$

so (3.3) is indeed valid.

Now we show that for any odd integers b, c which are coprime and greater than 1, we have

$$\sum_{i=1}^{\frac{c-1}{2}}\left\lfloor\frac{ib}{c}\right\rfloor+\sum_{j=1}^{\frac{b-1}{2}}\left\lfloor\frac{jc}{b}\right\rfloor=\frac{b-1}{2}\cdot\frac{c-1}{2}. \tag{3.4}$$

For this, consider the points

$$A=(0,0),\quad B=\left(\frac{b}{2},0\right),\quad C=\left(0,\frac{c}{2}\right),\quad D=\left(\frac{b}{2},\frac{c}{2}\right)$$

on the plain, and observe that the right hand side of (3.4) is just the number of points with integral coordinates in the *interior* of the rectangle T spanned by these points. (So we do not take into consideration the points on the sides of T.) We prove that the left hand side of (3.4) gives the number of these points, as well. For this, cut the rectangle T into two parts along the line AC, which is given by the formula

$$y=\frac{c}{b}x.$$

Observe that by $\gcd(b,c)=1$ the diagonal does not contain points with integral coordinates. First we count the integer points inside the triangle ABC; write P for this number. For this we count the integer points on the 'vertical' lines $x=j$ $(j=1,\dots,(b-1)/2)$ in the triangle, then we add up these numbers. The x coordinates of these points is obviously j, while their y coordinates satisfy $1\le y<\frac{c}{b}j$. So the number of these points is given by $\lfloor\frac{cj}{b}\rfloor$, whence

$$P=\sum_{j=1}^{\frac{b-1}{2}}\left\lfloor\frac{cj}{b}\right\rfloor.$$

A similar reasoning for the number Q of the integer points inside the triangle ACD, counting the points on the 'horizontal' lines $y = i$, yields

$$Q = \sum_{i=1}^{\frac{c-1}{2}} \left\lfloor \frac{bi}{c} \right\rfloor .$$

Since the number of integer points inside T is $P + Q$, the assertion (3.4) is indeed valid.

From the above verified two assertions our statement already follows. Indeed, (3.3) yields that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^w$$

holds, where

$$w = \sum_{i=1}^{\frac{p-1}{2}} \left\lfloor \frac{qi}{p} \right\rfloor + \sum_{j=1}^{\frac{q-1}{2}} \left\lfloor \frac{pj}{q} \right\rfloor ,$$

and here by (3.4) we have

$$w = \frac{p-1}{2} \cdot \frac{q-1}{2}.$$

After rearrangement, this implies our statement. □

3.3 The Jacobi symbol

In this section, as a generalization of the Legendre symbol, we study the Jacobi symbol and some of its properties.

Definition 3.3.1 *Let n be an odd integer greater than 1 and let $a \in \mathbb{Z}$. Then the Jacobi symbol $\left(\frac{a}{n}\right)$ is defined as*

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)^{\alpha_1} \dots \left(\frac{a}{p_r}\right)^{\alpha_r},$$

where $n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$ is the prime factorization of n.

Remark 3.3.1 Clearly, the Jacobi symbol is a generalization of the Legendre symbol. On the other hand, the Jacobi symbol is not directly related to quadratic residues modulo n:

$$\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = 1,$$

however, the congruence

$$x^2 \equiv 2 \pmod{15}$$

is not solvable in integers x.

The following theorems yield extensions of the corresponding ones concerning the Legendre symbol, to the case of the Jacobi symbol. We start with the extension of the fundamental properties.

Theorem 3.3.1 *Let $n \in \mathbb{N}$ be odd, $n > 1$. Then*

a) for any $a, b \in \mathbb{Z}$, $a \equiv b \pmod{n}$ implies that

$$\left(\frac{a}{n}\right) = \left(\frac{b}{n}\right),$$

b) for any $a, b \in \mathbb{Z}$ we have

$$\left(\frac{ab}{n}\right) = \left(\frac{a}{n}\right)\left(\frac{b}{n}\right),$$

c) for any $a, b \in \mathbb{Z}$, $\gcd(b, n) = 1$ implies that

$$\left(\frac{ab^2}{n}\right) = \left(\frac{a}{n}\right).$$

Proof. We prove the statements separately.

a) The assertion is a trivial consequence of the Jacobi symbol and the corresponding property of the Legendre symbol.

b) The claim follows from the (multplicativity preserving) definition of the Jacobi symbol and the multiplicative property of the Legendre symbol.

c) This assertion also follows easily from the definition of the Jacobi symbol and the corresponding property of the Legendre symbol. □

Our next theorem gives the Jacobi symbol $\left(\frac{a}{n}\right)$ for certain particular values of a.

Theorem 3.3.2 *Let $n \in \mathbb{N}$ be odd, $n > 1$. Then we have*

$$\left(\frac{1}{n}\right) = 1, \quad \left(\frac{-1}{n}\right) = (-1)^{\frac{n-1}{2}}, \quad \left(\frac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}}.$$

Proof. The first statement directly follows from the definition of the Jacobi symbol and the fact that

$$\left(\frac{1}{p}\right) = 1$$

for any prime p.

Let $p > 2$ prime. Then

$$\left(\frac{-1}{p}\right) = \begin{cases} 1, & \text{if } p \text{ is of the form } p = 4k+1, \\ -1, & \text{if } p \text{ is of the form } p = 4k-1. \end{cases}$$

So if $n > 1$ is odd and

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

is the prime factorization of n, then by

$$\left(\frac{-1}{n}\right) = \left(\frac{-1}{p_1}\right)^{\alpha_1} \dots \left(\frac{-1}{p_r}\right)^{\alpha_r}$$

the value of $\left(\frac{-1}{n}\right)$ is -1 if and only if the sum of the exponents of the primes of the form $4k-1$ in the prime factorization of n is odd. However, this precisely means that n is of the form $n = 4\ell - 1$. In other words,

$$\left(\frac{-1}{n}\right) = \begin{cases} 1, & \text{if } n \text{ is of the form } n = 4\ell+1, \\ -1, & \text{if } n \text{ is of the form } n = 4\ell-1. \end{cases}$$

From this our claim immediately follows.

We follow a similar path to prove our third assertion. Let $n > 1$ be odd, and

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

be the prime factorization of n again. Then by

$$\left(\frac{2}{n}\right) = \left(\frac{2}{p_1}\right)^{\alpha_1} \dots \left(\frac{2}{p_r}\right)^{\alpha_r},$$

using our knowledge concerning the Legendre symbol $\left(\frac{2}{p}\right)$ (p prime) we obtain that the value of $\left(\frac{2}{n}\right)$ is -1 if and only if the total sum of the exponents of the primes of the shape $8k-3$ and $8k+3$ in the prime factorization of n is odd. This precisely means that n is of the shape $n = 8\ell \pm 3$. In other words,

$$\left(\frac{2}{n}\right) = \begin{cases} 1, & \text{if } n \text{ is of the form } n = 8\ell \pm 1, \\ -1, & \text{if } n \text{ is of the form } n = 8\ell \pm 3. \end{cases}$$

From this our claim immediately follows. □

The last theorem of the section is the quadratic reciprocity law concerning the Jacobi symbol.

Theorem 3.3.3 (Quadratic reciprocity) *Let $m > 1$ and $n > 1$ be odd. Then we have*

$$\left(\frac{n}{m}\right) = (-1)^{\frac{(n-1)(m-1)}{4}} \left(\frac{m}{n}\right).$$

Proof. Let the prime factorization of n be given by

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r},$$

and the prime factorization of m be given by

$$m = q_1^{\beta_1} \dots q_t^{\beta_t}.$$

Then by the definition of the Jacobi symbol and the properties of the Legendre symbol we obtain

$$\left(\frac{m}{n}\right) = \prod_{i=1}^{r} \prod_{j=1}^{t} \left(\frac{q_j}{p_i}\right)^{\alpha_i \beta_j}$$

and

$$\left(\frac{n}{m}\right) = \prod_{i=1}^{r} \prod_{j=1}^{t} \left(\frac{p_i}{q_j}\right)^{\alpha_i \beta_j}.$$

Let

$$U = \{i : p_i \text{ is of the form } p_i = 4k-1 \text{ and } \alpha_i \text{ is even } (1 \le i \le r)\}$$

and

$$V = \{j : q_j \text{ is of the form } q_j = 4k-1 \text{ and } \beta_j \text{ is even } (1 \le j \le t)\}.$$

Then by the quadratic reciprocity law concerning the Legendre symbol we get

$$\left(\frac{p_i}{q_j}\right)^{\alpha_i\beta_j} = -\left(\frac{q_j}{p_i}\right)^{\alpha_i\beta_j} \quad \text{if } i \in U \text{ and } j \in V,$$

however,

$$\left(\frac{p_i}{q_j}\right)^{\alpha_i\beta_j} = \left(\frac{q_j}{p_i}\right)^{\alpha_i\beta_j} \quad \text{if } i \notin U \text{ or } j \notin V$$

holds. Hence

$$\left(\frac{m}{n}\right) = (-1)^{uv}\left(\frac{n}{m}\right),$$

where $u = |U|$, $v = |V|$. On the other hand, observe that uv is odd if and only if both u and v are odd, that is, if and only if n and m are both of the shape $4\ell - 1$. so we get that

$$\left(\frac{m}{n}\right) = \begin{cases} -\left(\frac{n}{m}\right), & \text{if both } n \text{ and } m \text{ are of the form } 4\ell-1, \\ \left(\frac{n}{m}\right), & \text{otherwise.} \end{cases}$$

From this the theorem immediately follows. □

3.4 Exercises

Exercise 3.4.1 Find the integer solutions x of the following quadratic congruences:

a) $x^2 \equiv 2 \pmod 5$

b) $x^2 \equiv 5 \pmod 7$

c) $x^2 \equiv 3 \pmod{11}$

d) $x^2 + 2x + 1 \equiv 3 \pmod{11}$

e) $x^2 - 1 \equiv 10 \pmod{11}$

f) $x^2 + 4x + 1 \equiv 3 \pmod{13}$

g) $x^2 \equiv 21 \pmod{17}$

h) $x^2 - 3x + 3 \equiv 3 \pmod{17}$

i) $x^2 - 1 \equiv 3 \pmod{19}$

j) $x^2 + 1 \equiv 3 \pmod{19}$

Exercise 3.4.2 Find those integers a for which the congruence

$$x^2 + 3x + a \equiv 0 \pmod{11}$$

is solvable!

Exercise 3.4.3 Let p be an odd prime, a, b, c integers. Prove that if abc, ab, ac, bc are all quadratic residues modulo p, then a, b and c are also quadratic residues!

Exercise 3.4.4 Let p be an odd prime, a, b, c integers. Prove that if ab, ac, bc are all quadratic residues modulo p, then either a, b and c are all quadratic residues, or they are all quadratic non-residues!

Exercise 3.4.5 Let p be an odd prime, $a \in \mathbb{Z}$ and $k \in \mathbb{N}$. Prove that then

$$\left(\frac{a}{p}\right) = \left(\frac{a^{2k+1}}{p}\right)$$

holds!

Exercise 3.4.6 Let p be an odd prime, $a, b \in \mathbb{Z}$ such that

$$ab \equiv 1 \pmod{p}.$$

Prove that then

$$\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$$

holds!

Exercise 3.4.7 Let p be an odd prime, and $a \in \mathbb{Z}$ a generator modulo p. Prove that then

$$\left(\frac{a}{p}\right) = -1$$

holds!

Exercise 3.4.8 Let p be an odd prime, and $a, b \in \mathbb{Z}$ be generators modulo p. Prove that then ab is not a generator modulo p.

Exercise 3.4.9 Prove that if p is a prime of the form $p = 4k - 1$, then the congruence

$$2x^2 - 2x + 1 \equiv 0 \pmod{p}$$

is not solvable!

Exercise 3.4.10 For which primes p is the congruence

$$x^2 + 2 \equiv 0 \pmod{p}$$

solvable?

Exercise 3.4.11 Prove that for any integer n the odd divisors of $n^2 + 1$ are of the form $4k + 1$.

Exercise 3.4.12 Let p be a prime of the form $p = 4k + 1$. Calculate the values of the sums

$$\left(\frac{1}{p}\right) + \left(\frac{2}{p}\right) + \ldots + \left(\frac{(p-1)/2}{p}\right)$$

and

$$\left(\frac{1}{p}\right) + \left(\frac{3}{p}\right) + \left(\frac{5}{p}\right) + \ldots + \left(\frac{p-2}{p}\right).$$

Exercise 3.4.13 Prove that if p is a prime of the form $p = 2^n + 1$ with $n \geq 1$ (i.e. p is a so-called Fermat prime), then $a \in \mathbb{Z}$ is a generator modulo p if and only if

$$\left(\frac{a}{p}\right) = -1$$

holds!

Exercise 3.4.14 Prove that if p is a prime of the form $p = 2^n + 1$ with $n > 1$ (i.e. p is a Fermat prime again), then 3 is a generator modulo p.

Exercise 3.4.15 Describe those primes p for which the following congruences are solvable:

a) $x^2 \equiv 3 \pmod{p}$

b) $x^2 \equiv 5 \pmod{p}$

c) $x^2 \equiv 13 \pmod{p}$

Exercise 3.4.16 Prove that

$$2^{37 \cdot 73 - 1} \equiv 1 \pmod{37 \cdot 73}$$

holds!

Exercise 3.4.17 What is the remainder if we divide 106^{106} by 211 ?

Exercise 3.4.18 Calculate the values of the following Legendre symbols:

a) $\left(\frac{9}{37}\right)$

b) $\left(\frac{-9}{37}\right)$

c) $\left(\frac{101}{103}\right)$

d) $\left(\frac{50}{101}\right)$

e) $\left(\frac{33}{89}\right)$

f) $\left(\frac{32}{89}\right)$

g) $\left(\frac{17}{71}\right)$

h) $\left(\frac{128}{11}\right)$

i) $\left(\frac{512}{13}\right)$

j) $\left(\frac{-512}{13}\right)$

Exercise 3.4.19 Calculate the values of the following Jacobi symbols:

a) $\left(\frac{9}{15}\right)$

b) $\left(\frac{-9}{15}\right)$

c) $\left(\frac{11}{39}\right)$

d) $\left(\frac{26}{27}\right)$

e) $\left(\frac{33}{25}\right)$

f) $\left(\frac{32}{33}\right)$

g) $\left(\frac{17}{51}\right)$

h) $\left(\frac{128}{55}\right)$

i) $\left(\frac{512}{55}\right)$

j) $\left(\frac{-512}{55}\right)$

Chapter 4

Elements of prime number theory

In this chapter we discuss some notions and assertions from the field of prime number theory. We mostly concentrate on those points which are related to the objects studied before, or prove to be useful later on. Beside these we study certain other things which are of general interest.

4.1 Problems concerning primes

In this section we study both elementary, basic assertions and classical problems. The first theorem is long and well-known, its proof goes back to Euclid.

Theorem 4.1.1 *There are infinitely many primes.*

Proof. Suppose to the contrary that there are only finitely many primes, given by $p_1, \ldots, p_n$. Consider the number

$$A := p_1 \ldots p_n + 1.$$

According to the fundamental theorem of arithmetic A has a prime divisor p. Since $\gcd(A, p_i) = 1$, we have $p \neq p_i$ $(i = 1, \ldots, n)$. This contradiction proves our claim. □

Remark 4.1.1 By the sieve of Eratosthenes in principle all the primes can be given. List the integers greater than 1, and rule out all which are divisible

by 2, then by 3, and so on; in general, rule out those which are divisible by the smallest number not ruled out:

$$2, 3, \not{4}, 5, \not{6}, 7, \not{8}, \not{9}, \ldots$$

As the result of the procedure we obtain the prime numbers.

Definition 4.1.1 *Write p_n for the n-th prime; that is, $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc.*

The next theorem gives a simple, however (as we shall see later) rather rough upper bound for the magnitude of the n-th prime.

Theorem 4.1.2 *For every $n \in \mathbb{N}$ we have $p_n \leq 2^{2^{n-1}}$.*

Proof. We prove the theorem by induction. If $n = 1$, then the formula gives $2 \leq 2$, which is true. Assume that the statement is valid for all $k \leq n$ with some positive integer n. Then we have

$$A := p_1 \ldots p_n + 1 \leq 2^{1+2+\ldots+2^{n-1}} + 1 \leq 2^{2^n-1} + 2^{2^n-1} = 2^{2^n}.$$

Thus A has a prime divisor p such that $p \leq 2^{2^n}$. As clearly $p \neq p_i$ $(i = 1, \ldots, n)$, hence

$$p_{n+1} \leq p \leq 2^{2^n},$$

which proves our statement. □

Now we introduce two famous families of prime numbers.

Definition 4.1.2 *The primes of the form $2^m - 1$ are called Mersenne primes, and the primes of the form $2^m + 1$ are called Fermat primes.*

Theorem 4.1.3 *If $2^m - 1$ is a Mersenne prime, then m is a prime. If $2^m + 1$ is a Fermat prime, then m is a power of 2.*

Proof. Suppose first that $2^m - 1$ is a Mersenne prime. Assume to the contrary that m is not a prime. Then we can write $m = ab$ with some integers a, b satisfying $1 < a \leq b < m$. Thus

$$2^m - 1 = 2^{ab} - 1 = (2^a)^b - 1^b = (2^a - 1)(2^{a(b-1)} + 2^{a(b-2)} + \cdots + 1),$$

which, since both factors on the right hand side are greater than 1, yields a contradiction. Thus m must be a prime, indeed.

Let now 2^m+1 be a Fermat prime. Assume to the contrary that m is not a power of 2. Then we have $m = ab$ with some integers $1 < a, b < m$, where a is odd. Therefore

$$2^m + 1 = 2^{ab} + 1 = (2^b)^a + 1^a = (2^b + 1)(2^{b(a-1)} - 2^{b(a-2)} + \cdots + 1),$$

which, since both factors on the right hand side are greater than 1 again, yields a contradiction. So m is necessarily a power of 2. $\square$

Remark 4.1.2 According to standard conjectures, there are infinitely many Mersenne primes, however, the number of Fermat primes is finite. It is clear that not all numbers of the form $2^p - 1$ (p prime) or of the shape $2^{2^n} + 1$ ($n \geq 0$) are primes. For example, we have

$$2^{11} - 1 = 2047 = 23 \cdot 89$$

and

$$2^{2^5} + 1 = 4294967297 = 641 \cdot 6700417.$$

(The latter factorization was already known to Euler.)

The next theorem establishes an important link between Mersenne primes and perfect numbers.

Theorem 4.1.4 *An even positive integer n is a perfect number if and only if it is of the form*

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

where $2^m - 1$ is a Mersenne prime.

Proof. Let $2^m - 1$ be a Mersenne prime, and put

$$n = 2^{m-1}(2^m - 1).$$

Then by $\gcd(2^{m-1}, 2^m - 1) = 1$ and the multiplicativity of σ we obtain

$$\sigma(n) = \sigma(2^{m-1}(2^m - 1)) = \sigma(2^{m-1})\sigma(2^m - 1) = \\ = \frac{2^m - 1}{2 - 1} \cdot (2^m - 1 + 1) = 2^m(2^m - 1) = 2n.$$

Thus the numbers of the form given in the statement are perfect numbers, indeed.

Assume now that n is an even perfect number. Then we can write

$$n = 2^a \cdot b$$

with $a \geq 1$ and b odd. As n is a perfect number, thus

$$\sigma(n) = \sigma(2^a \cdot b) = (2^{a+1} - 1)\sigma(b) = 2n.$$

Hence we get

$$(2^{a+1} - 1)\sigma(b) = 2^{a+1}b,$$

which can be rewritten as

$$(2^{a+1} - 1)(\sigma(b) - b) = b. \tag{4.1}$$

Thus in particular

$$\sigma(b) - b \mid b.$$

Since $a \geq 1$, we also have

$$\sigma(b) - b < b.$$

That is, $\sigma(b) - b$ and b are two distinct divisors of b. Therefore by

$$\sigma(b) = (\sigma(b) - b) + b$$

b cannot have more divisors. This shows that b is a prime, whence $\sigma(b) = b + 1$. Further, (4.1) yields that

$$2^{a+1} - 1 = b$$

is also valid. So n is of the form

$$n = 2^a(2^{a+1} - 1)$$

where $2^{a+1} - 1$ is a prime, which was to be proved. □

As a simple consequence of the above theorem we obtain the following statement.

Corollary 4.1.1 *There exist infinitely many even perfect numbers if and only if there exist infinitely many Mersenne primes.*

Remark 4.1.3 Goldbach in 1742 formulated the following conjectures.

Even Goldbach conjecture. Every even integer greater than 2 can be expressed as the sum of two primes.

Odd Goldbach conjecture. Every odd integer greater than 5 can be expressed as the sum of three primes.

The first problem appeared among the Millennium problems (of prize money one million USD each) of the Clay Mathematics Institute. The even Goldbach conjecture implies the odd one: if n is odd and $n-3 = p_1 + p_2$, then certainly $n = p_1 + p_2 + 3$.

Vinogradov in 1937 showed that there exists an n_0 such that if n is odd and $n > n_0$, then n can be written as the sum of three primes. Borodzkin in 1956 proved that n_0 can be taken as $3^{14348907}$, while Chen and Vang in 1989 showed that $n_0 = 10^{43000}$ is also an appropriate choice. Finally, in 2015 Helfgott proved the odd Goldbach conjecture.

Definition 4.1.3 *If p and q are primes and their difference is 2, then we say that p and q are twin primes.*

Remark 4.1.4 According to a conjecture based upon the common opinion of the experts of the topic, there are infinitely many twin primes. For example, $3, 5$; $5, 7$; $11, 13$ are twin primes. By our present knowledge the solution of the twin prime conjecture cannot be expected in the near future. On the other hand, rather recently, by the pioneering work of Goldston, Pintz, Yildirim, Zhang, Tao and others we know that there exist infinitely many primes p, q such that $|p - q| \leq 246$.

The following theorems are about the relation of primes and arithmetic progressions. The first results of this type - which we give without a proof - is a relatively fresh result (from 2004) due to Green and Tao, which answers an ancient problem.

Theorem 4.1.5 (Green-Tao) *The set of primes contains arbitrary long finite arithmetic progressions.*

Now we formulate a famous theorem of Dirichlet, concerning primes in arithmetic progressions.

Theorem 4.1.6 (Dirichlet) *Let $a \in \mathbb{N}$, $b \in \mathbb{Z}$, $\gcd(a, b) = 1$. Then the arithmetic progression $ax + b$ ($x \in \mathbb{N}$) contains infinitely many primes.*

Proof. The proof of the theorem requires rather deep tools, which are beyond the scope of the present lecture notes. So we restrict or discussion only to a few special cases. Beforehand we note that the cases $a = 1$ and $a = 2$ are trivial.

Let $(a, b) = (4, -1)$. Assume to the contrary that there exist only finitely many primes of the form $4k - 1$, given by $q_1, \ldots, q_n$. Further, put

$$A = 4q_1 \ldots q_n - 1.$$

Then A must clearly have a prime divisor of the form $4k - 1$, which is obviously different from $q_1, \ldots, q_n$. This contradiction implies the theorem in this case.

Let now $(a, b) = (4, 1)$. Suppose to the contrary that there are only finitely many primes of the shape $4k + 1$, given by $q_1, \ldots, q_n$. Set now

$$A = (2q_1 \ldots q_n)^2 + 1,$$

and let p be a prime divisor of A. Then we clearly have $p \neq 2$, furthermore, $p \neq q_i$ $(i = 1, \ldots, n)$. On the other hand, by

$$(2q_1 \ldots q_n)^2 \equiv -1 \pmod{p}$$

we have

$$\left(\frac{-1}{p}\right) = 1.$$

However, then Theorem 3.2.2 implies that

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}},$$

whence

$$(-1)^{\frac{p-1}{2}} = 1.$$

Thus $\frac{p-1}{2}$ is odd, so the prime p is of the form $4k+1$. This is a contradiction, and the statement follows also in this case.

Let $(a, b) = (6, -1)$. Assume that there are only finitely many primes of the form $6k - 1$, which are $q_1, \ldots, q_n$. Further, write

$$A = 6q_1 \ldots q_n - 1.$$

Then obviously A must have a prime divisor of the shape $6k-1$, which is different from all the primes $q_1, \ldots, q_n$. This contradiction implies our claim.

Finally, take $(a,b) = (6,1)$. Assume that there are only finitely many primes of the form $6k+1$, which are given by $q_1, \ldots, q_n$. Let now

$$A = (2q_1 \ldots q_n)^2 + 3,$$

and let p be a prime divisor of A. Then clearly $p \neq 2$, $p \neq 3$ and also $p \neq q_i$ $(i = 1, \ldots, n)$. On the other hand,

$$(2q_1 \ldots q_n)^2 \equiv -3 \pmod{p}$$

implies that

$$\left(\frac{-3}{p}\right) = 1$$

holds. However, the properties of the Legendre symbol imply that

$$\left(\frac{-3}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{3}{p}\right),$$

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$$

and

$$\left(\frac{3}{p}\right) = (-1)^{\frac{2(p-1)}{4}}\left(\frac{p}{3}\right).$$

Thus

$$(-1)^{\frac{p-1}{2}}(-1)^{\frac{p-1}{2}}\left(\frac{p}{3}\right) = 1,$$

whence we get

$$\left(\frac{p}{3}\right) = 1.$$

As

$$\left(\frac{-1}{3}\right) = -1,$$

p is necessarily of the form $6k+1$. This is a contradiction which proves the statement also in this case. $\square$

4.2 On the distribution of primes

In this section we discuss theorems concerning the distribution of the primes.

Definition 4.2.1 *Let* $\Pi(x) = \sum_{p \leq x} 1$*, the number of primes not greater than* x*.*

Remark 4.2.1 One can easily check that $\Pi(x)$ is a non-negative step function which is continuous from the right, $\Pi(2) = 1$, $\Pi(3) = 2$, $\Pi(5) = 3$, etc.

The following famous theorem is due to Chebishev.

Theorem 4.2.1 (Chebishev) *There exists real numbers* c_1, c_2 *with* $0 < c_1 < 1 < c_2$ *such that for any* $x \geq 2$

$$c_1 \frac{x}{\log x} < \Pi(x) < c_2 \frac{x}{\log x}$$

holds.

To formulate the theorem describing the asymptotic behavior of the function $\Pi(x)$, we need the following notion.

Definition 4.2.2 *Let* $f, g : \mathbb{R} \to \mathbb{R}$ *be two functions. If*

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1,$$

then we say that f *and* g *are asymptotically equal. Notation:* $f \sim g$*.*

The next theorem - given without a proof - is called the large prime number theorem. It was first proved independently by Hadamard and de la Vallée Poussin in 1896.

Theorem 4.2.2 (Large prime number theorem) *We have*

$$\Pi(x) \sim \frac{x}{\log x}.$$

The large prime number theorem shows that the set of primes is 'rare'. For the precise formulation of this property we need the following concept.

Definition 4.2.3 *Let $(a_n)_{n=1}^{\infty}$ be a strictly monotone increasing sequence of positive integers. Then by the density of this sequence we mean the quantity*

$$S((a_n)_{n=1}^{\infty}) = \lim_{x\to\infty} \frac{\#\{a_n : n \in \mathbb{N},\ a_n \leq x\}}{x},$$

provided that the limit exists.

The large prime number theorem immediately implies the following statement.

Corollary 4.2.1 *The sequence of primes is of density zero.*

Proof. The large prime number theorem implies that

$$\frac{\Pi(x)}{x} \sim \frac{1}{\log x},$$

whence

$$\lim_{x\to\infty} \frac{\Pi(x)}{x} = 0.$$

This proves our claim. □

The next theorem - which we give without a proof - is known as Bertrand's postulate.

Theorem 4.2.3 *For every $n \in \mathbb{N}$, $n > 1$ the interval $(n, 2n)$ contains a prime.*

The above theorem is valid in the following form - which is given without a proof again - as well.

Theorem 4.2.4 *For every positive real number δ there exists a positive real number x_0, such that for any $x > x_0$ the interval $(x, (1+\delta)x)$ contains a prime.*

By the help of the large prime number theorem one can get a fairly good estimate for the n-th prime, as well.

Theorem 4.2.5 *For the n-th prime we have $p_n \sim n \log n$, that is*

$$\lim_{n\to\infty} \frac{p_n}{n \log n} = 1.$$

Remark 4.2.2 The above three theorems are in fact simple consequences of the large prime number theorem (more precisely, in case of Bertrand's postulate we need a quantitative version of it). The details are left to the Reader.

Our next theorem shows that $\mathbb{N}$ contains arbitrary long intervals of composite numbers.

Theorem 4.2.6 *There exists arbitrary long intervals which do not contain primes.*

Proof. For any $n \in \mathbb{N}$, none of the numbers

$$(n+1)! + 2, \ldots, (n+1)! + (n+1)$$

is a prime, which proves our claim. □

Remark 4.2.3 Observe that as the magnitude of $n!$ is around $\left(\frac{n}{e}\right)^n$, the above interval consists of rather large numbers compared to its length.

Our next theorem is a classical result of prime number theory. We give the proof due to Euler.

Theorem 4.2.7 *The series $\sum_p \frac{1}{p}$ is divergent.*

Proof. We prove more than what we claimed: we show that for any $x > 1$

$$\sum_{p \le x} \frac{1}{p} > \log\log x - 2$$

holds. For this we shall use the following well-known assertions from analysis:
a) $\sum_{k \le x} \frac{1}{k} > \log x$ if $x > 1$,
b) $\log \frac{1}{1-x} = x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots \le x + x^2$ if $0 \le x \le \frac{1}{2}$,

c) $\sum\limits_{k\le x}\frac{1}{k^2}<2$ if $x>1$.

Consider the product

$$A_x=\prod_{p\le x}\left(1+\frac{1}{p}+\ldots+\frac{1}{p^{\nu_p}}\right),$$

where $x>1$ and

$$p^{\nu_p-1}\le x<p^{\nu_p}.$$

Obviously we have

$$A_x\ge\sum_{k\le x}\frac{1}{k},$$

since k has a prime factorization of the form

$$k=p_1^{\alpha_1}\ldots p_r^{\alpha_r},$$

where $p_i\le x$ and $\alpha_i\le\nu_{p_i}$. Thus by a) we obtain $A_x>\log x$. On the other hand, as the terms of A_x are sums of consecutive terms of geometric series, we get that

$$A_x=\prod_{p\le x}\frac{\left(\frac{1}{p}\right)^{\nu_p+1}-1}{\frac{1}{p}-1}<\prod_{p\le x}\frac{1}{1-\frac{1}{p}}.$$

Therefore

$$\prod_{p\le x}\frac{1}{1-\frac{1}{p}}>\log x.$$

Taking logarithm we get

$$\sum_{p\le x}\log\frac{1}{1-\frac{1}{p}}>\log\log x,$$

which by b) implies that

$$\sum_{p\le x}\left(\frac{1}{p}+\frac{1}{p^2}\right)>\log\log x.$$

Finally, c) yields that

$$\sum_{p\le x}\frac{1}{p}>\log\log x-2.$$

This assertion implies the theorem. □

Remark 4.2.4 It is well-known that

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

So if $(a_n)_{n=1}^{\infty}$ is a strictly monotone increasing sequence in $\mathbb{N}$, then by the sum of the series

$$\sum_{n=1}^{\infty} \frac{1}{a_n}$$

we can measure how quickly the sequence increases. Thus the previous theorem shows that p_n increases relatively slowly, in other words, the primes are situated relatively dense. On the other hand, we have

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6},$$

so the sequence of squares increases relatively quickly. This motivates the following conjecture:

Conjecture. There is a prime between any two squares.

By Bertrand's postulate we know that the interval $(n^2, 2n^2)$ (for $n > 1$) contains a prime. To prove the above conjecture we should show that already the interval $(n^2, n^2 + 2n + 1)$ contains a prime.

4.3 Exercises

Exercise 4.3.1 Prove that if $n > 2$, then $2^n - 1$ and $2^n + 1$ cannot be twin primes!

Exercise 4.3.2 Prove that for $k > 3$ it cannot happen that both p_{k-1}, p_k and p_k, p_{k+1} are twin primes! (Recall that p_n denotes the n-th prime.)

Exercise 4.3.3 Prove that there are infinitely many primes which are not members of any pair of twin primes!

Exercise 4.3.4 Prove that any positive integer greater than 11 can be represented as the sum of two composite numbers!

Exercise 4.3.5 Prove that if $n > 1$, then the Fermat number $2^{2^n} + 1$ is not the sum of two primes!

Exercise 4.3.6 Prove that there are infinitely many odd positive integers which cannot be written as the sum of two primes!

Exercise 4.3.7 Calculate $\Pi(8)$, $\Pi(11)$ and $\Pi(20)$.

Exercise 4.3.8 Decide whether the function $\Pi(n)$ $(n \in \mathbb{N})$ is multiplicative or additive!

Exercise 4.3.9 Prove that a positive integer n with $n > 1$ is a prime if and only if

$$\frac{\Pi(n-1)}{n-1} < \frac{\Pi(n)}{n}$$

holds!

Exercise 4.3.10 Prove that if $x, y \in \mathbb{N}$ with $x > 1$ and $2 \leq y \leq 10$, then

$$\Pi(x+y) \leq \Pi(x) + \Pi(y)$$

holds!

Exercise 4.3.11 Decide which of the following series are convergent:

$$\sum_{p \text{ prime}} \frac{1}{p+2}, \qquad \sum_{\substack{p,q \text{ prime} \\ p \neq q}} \frac{1}{pq}, \qquad \sum_{p \text{ prime}} \frac{1}{p^2}.$$

Exercise 4.3.12 Let d_n denote the n-th prime difference, that is let

$$d_n = p_{n+1} - p_n \quad (n \geq 1).$$

(Here p_n is the n-th prime.) Prove that for infinitely many $k \in \mathbb{N}$

$$d_{k+1} > d_k$$

holds!

Exercise 4.3.13 Let d_n be as in the previous exercise. Prove that for any $n \in \mathbb{N}$

$$n - 1 \leq \sum_{k=1}^{\Pi(n)} d_k \leq 2n - 2$$

holds!

Exercise 4.3.14 Let $a \in \mathbb{N}$, $a > 1$ be arbitrary. Prove that there exist infinitely many primes whose first digit in the base a number system is 1.

Exercise 4.3.15 Prove that for $n > 1$ the number $n!$ is not the power (with exponent greater than 1) of some positive integer!

Exercise 4.3.16 Let $n \in \mathbb{N}$. Prove that at least one of the numbers n and $n + 1$ can be expressed as the sum of different primes!

Exercise 4.3.17 Let $a, m \in \mathbb{N}$, $b \in \mathbb{Z}$, $\gcd(a, b) = \gcd(m, b) = 1$. Prove that there exist infinitely many $x \in \mathbb{N}$, for which $m \mid x$ and $ax+b$ is a prime!

Exercise 4.3.18 Prove that there are infinitely many primes of the form

$$6 + 10x + 15y \quad (x \geq 0,\ y \geq 0).$$

Exercise 4.3.19 Let $a, n \in \mathbb{N}$, $b \in \mathbb{Z}$, $\gcd(a, b) = 1$. Prove that the arithmetic progression $ax + b$ $(x = 0, 1, 2, \dots)$ has infinitely many terms which are the products of n distinct primes!

Exercise 4.3.20 Prove that for any $k \in \mathbb{N}$ one can find a prime p, such that in the base 10 representation of p there are at least k zero digits!

Exercise 4.3.21 Prove that for any $k \in \mathbb{N}$ one can find a prime p, such that in the base 10 representation of p the sum of the digits is at least k.

Exercise 4.3.22 Prove that for any $b \in \mathbb{N}$ there exist infinitely many $n \in \mathbb{N}$, such that $d(n^b)$ is a prime!

Exercise 4.3.23 Prove that for any odd $a \in \mathbb{N}$ there exist infinitely many odd $n \in \mathbb{N}$, such that $\gcd(a, \varphi(n)) = 1$.

Exercise 4.3.24 Prove that for any $a \in \mathbb{N}$ and prime p there exist infinitely many primes q, for which

$$p \mid a^q - a$$

holds!

Exercise 4.3.25 Prove that the number

$$0.p_1p_2p_3p_4p_5p_6\ldots = 0.23571113\ldots$$

is irrational!

Exercise 4.3.26 Let $n \in \mathbb{N}$ be not a square. Prove that then for infinitely many prime p the number n is a quadratic non-residue modulo p.

Chapter 5

Pseudoprimes and probabilistic prime tests

In this chapter we study probabilistic prime tests. For this, we shall need to discuss pseudoprimes and their properties, as well.

5.1 Some basic notions of complexity

In this section we introduce some basic notions of complexity theory.

Definition 5.1.1 *The following operations appearing during the addition of two binary numbers a and b are called bit operations:*

- *reading the next bits of a and b and checking whether there was a carry,*
- *if both bits are 0 and there is no carry, then writing a 0,*
- *if both bits are 0 and there is a carry, or if precisely on of the bits is 0 and there is no carry, then writing a 1,*
- *if both bits are 1 and there is no carry, or if precisely on of the bits is 1 and there is a carry then writing a 0 and noting that we have a carry,*
- *if both bits are 1 and there is a carry then writing a 1 and noting that we have a carry.*

Definition 5.1.2 *Let $f, g : \mathbb{N}^r \to \mathbb{R}^+$. If there exists a positive real number c such that*

$$\frac{f(n_1, \ldots, n_r)}{g(n_1, \ldots, n_r)} < c$$

holds for any $(n_1, \ldots, n_r) \in \mathbb{N}^r$, then we write $f = O(g)$. If $r = 1$ and

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0,$$

then we write $f = o(g)$.

Remark 5.1.1 The magnitude of necessary bit operations for performing certain procedures are:

- adding two integers of k bits: k,
- multiplying two integers of k bits with the usual method: $O(k^2)$; with fast multiplication: $O(k^{\log_2 3})$,
- the calculation of $n!$: $O(n^2 \log^2 n)$,
- the Euclidean algorithm for the integers a, b: $O(\log^3 \max(a, b))$,
- calculation of $b^n \pmod m$: $O(\log n \log m^2)$.

For the latter one, an efficient procedure is the following. Let

$$n = \sum_{i=0}^{k-1} n_i 2^i$$

be the binary representation of n. Let $a = 1$, $b_0 = b$. Assume that we are after the i-th step. Let

$$a \equiv a b_{i+1}^{n_{i+1}} \pmod m$$

and let

$$b_{i+2} \equiv b_{i+1}^2 \pmod m$$

(in reduced form). We repeat this procedure while $i < k - 1$ holds. At the end we obtain

$$a \equiv b^n \pmod m$$

The next notion is of utmost importance: in general, polynomial algorithms are considered to be efficient.

Definition 5.1.3 *Let the input of an algorithm be given by $n_1, \ldots, n_r \in \mathbb{N}$. If there exist non-negative integers $d_1, \ldots, d_r$ such that for any input $n_1, \ldots, n_r$ the algorithm terminates after executing*

$$O((\log n_1)^{d_1} \ldots (\log n_r)^{d_r})$$

bit operations, then we say that algorithm is polynomial.

5.2 Pseudoprimes and Carmichael-numbers

We start the section with an interesting and important remark.

Remark 5.2.1 By the Euler-Fermat theorem we know that if p is a prime and $p \nmid a$, then

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

Chinese mathematicians in the middle ages thought that if $n > 2$ then n is a prime if and only if

$$2^{n-1} \equiv 1 \pmod{n}.$$

This is valid for $n < 341$, however, for $n = 11 \cdot 31 = 341$ the assertion fails.

Example 5.2.1 63 is not a prime, because

$$2^{62} \equiv 4 \cdot (2^6)^{10} \equiv 4 \pmod{63}.$$

Observe that though the above argument implies that 63 is composite, we have no idea about its prime factorization.

The above example offers a possibility to test whether a given integer is a prime or not. For this, we shall need the following notion.

Definition 5.2.1 *Let n be an odd composite number and b an integer such that $\gcd(n, b) = 1$ and*

$$b^{n-1} \equiv 1 \pmod{n}.$$

Then we say that n is a pseudoprime with respect to b.

Example 5.2.2 We have

$$3^{90} \equiv 1 \pmod{91},$$

however,

$$2^{90} \equiv 64 \not\equiv 1 \pmod{91}.$$

Thus 91 is a pseudoprime with respect to 3, but not with respect to 2.

The next definition in fact only recalls a well-known notation.

Definition 5.2.2 *If n is an odd positive integer and $b \in \mathbb{Z}$ with $\gcd(n, b) = 1$, then b^{-1} denotes an integer with*

$$b \cdot b^{-1} \equiv 1 \pmod{n}.$$

Our next theorem gives some important properties of pseudoprimes.

Theorem 5.2.1 *Let n be an odd positive integer. Then*

a) n is a pseudoprime with respect to b if and only if $\mathrm{ord}_n(b) \mid n-1$,

b) if n is a pseudoprime with respect to b_1 and b_2, then b is a pseudoprime with respect to $b_1 b_2$ and $b_1 b_2^{-1}$, as well,

c) if n is not a pseudoprime with respect to some b coprime to n, then n is not a pseudoprime with respect to at least half of the bases coprime to n, counted modulo n.

Proof. We prove our statements separately.

a) If

$$\mathrm{ord}_n(b) = d \mid n-1,$$

then clearly

$$b^{n-1} \equiv 1 \pmod{n}.$$

Assume not that we have

$$b^{n-1} \equiv 1 \pmod{n}.$$

Further, let $\mathrm{ord}_n(b) = d$, and

$$n - 1 = qd + r \quad (q, r \in \mathbb{Z}, 0 \leq r < n-1).$$

Then

$$b^r \equiv b^{n-1-qd} \equiv 1 \pmod{n},$$

whence we get $r = 0$ and $d \mid n-1$.

b) If

$$b_1^{n-1} \equiv 1 \pmod{n}$$

and

$$b_2^{n-1} \equiv 1 \pmod{n}$$

then after multiplication and taking inverse we obtain

$$(b_1 b_2)^{n-1} \equiv 1 \pmod{n}$$

and

$$(b_1 b_2^{-1})^{n-1} \equiv 1 \pmod{n}.$$

c) Let

$$1 = b_1 < \ldots < b_s \leq n-1$$

be all the bases for which n is a pseudoprime. Then n cannot be a pseudoprime with respect to any of the bases

$$b \cdot b_i \quad (i = 1, \ldots, s),$$

since otherwise we would get a contradiction by b). As in view of $\gcd(n, b) = 1$ the bases $b \cdot b_i$ are pairwise incongruent modulo n, thus the number of those residue classes which are coprime to n, and with respect to them n is not a pseudoprime, is at least s. □

The next theorem shows that pseudoprimality is not too rare.

Theorem 5.2.2 *There are infinitely many integers which are pseudoprime with respect to* 2.

Proof. We prove that if n is an odd pseudoprime with respect to 2, then $2^n - 1$ is also a pseudoprime with respect to 2. As 341 is a possible starting point, this obviously implies the statement.

So let n be an odd composite number for which

$$2^{n-1} \equiv 1 \pmod{n}.$$

Let

$$n = u \cdot v \quad (1 < u, v < n),$$

and put $m = 2^n - 1$. Observe that $2^u - 1 \mid m$, thus m is an odd composite number. Set

$$2^{n-1} - 1 = k \cdot n.$$

Then

$$2^{m-1} = 2^{2^n-2} = 2^{2kn}.$$

Therefore

$$m = 2^n - 1 \mid 2^{2kn} - 1 = 2^{m-1} - 1,$$

whence

$$2^{m-1} \equiv 1 \pmod{m}.$$

This proves our statement. □

The next definition is important: we introduce a rather nice, however, from some viewpoint 'dangerous' type of numbers.

Definition 5.2.3 *Let n be a composite number. If for all $b \in \mathbb{Z}$ with $\gcd(b, n) = 1$ we have*

$$b^{n-1} \equiv 1 \pmod{n},$$

then n is called a Carmichael-number.

Example 5.2.3 $561 = 3 \cdot 11 \cdot 17$ is a Carmichael-number.

Our next theorem - which we give without a proof - describes certain properties of Carmichael-numbers.

Theorem 5.2.3 *Let n be an odd composite number. Then*

a) if n is a Carmichael-number, then it is square-free,

b) if n is square-free, then n is a Carmichael-number if and only if $p - 1 \mid n - 1$ holds for every prime $p \mid n$.

The next theorem, due to Alford, Granville and Pomerance - which we also give without a proof - shows that there exist infinitely many Carmichael-numbers.

Theorem 5.2.4 *The number of Carmichael-numbers is infinite. Further, there exists an x_0, such that for any $x > x_0$ we have*

$$|\{y : y \text{ is a Carmichael-number}, \ y \leq x\}| > x^{2/7}.$$

Remark 5.2.2 Because of Carmichael-numbers, the Euler-Fermat theorem is not applicable directly for prime testing.

5.3 Euler pseudoprimes and the Solovay-Strassen prime test

In this section, by refining the pseudoprime property, we shall get a probability prime test.

Definition 5.3.1 *If n is an odd composite number, $b \in \mathbb{Z}$, $\gcd(n, b) = 1$ and*

$$b^{\frac{n-1}{2}} \equiv \left(\frac{b}{n}\right) \pmod{n},$$

then n is called an Euler-pseudoprime with respect to b.

Remark 5.3.1 If p is a prime and $p \nmid b$, then

$$b^{\frac{p-1}{2}} \equiv \left(\frac{b}{p}\right) \pmod{p}.$$

So primes satisfy the above congruence.

Our next theorem verifies that the new notion is a sharpening of pseudoprimes.

Theorem 5.3.1 *If n is an Euler-pseudoprime with respect to b, then n is a pseudoprime with respect to b.*

Proof. If

$$b^{\frac{n-1}{2}} \equiv \left(\frac{b}{n}\right) \pmod{n},$$

then after squaring we obtain

$$b^{n-1} \equiv 1 \pmod{n}.$$

This proves our claim. □

Now we show that for any n one can find 'many' bases b, such that n is *not* an Euler-pseudoprime with respect to b.

Theorem 5.3.2 *Let n be an odd composite number. Then n is not a pseudoprime with respect to at least half of the bases coprime to n, counted modulo n.*

Proof. First we show that there exists a $b \in \mathbb{Z}$ with $\gcd(n, b) = 1$, for which n is not an Euler-pseudoprime.

Assume first that n is not square-free and let p be a prime, for which $p^2 \mid n$. Put $b = 1 + n/p$. Then obviously $\gcd(n, b) = 1$. Further, by the definition of the Jacobi symbol we get

$$\left(\frac{b}{n}\right) = 1,$$

while

$$b^{\frac{n-1}{2}} = (1 + n/p)^{\frac{n-1}{2}} = \sum_{i=0}^{\frac{n-1}{2}} \binom{\frac{n-1}{2}}{i} (n/p)^i \equiv 1 + \frac{n-1}{2} \cdot \frac{n}{p} \not\equiv 1 \pmod{n}$$

also holds by $p \nmid n - 1$. Thus n is not an Euler-pseudoprime with respect to b.

Let now n be an arbitrary square-free integer and let p be a prime divisor of n. Note that $\gcd(p, n/p) = 1$ and $n/p > 1$. Let a be a quadratic non-residue modulo p, and consider the following linear congruence system:

$$\begin{cases} x \equiv a \pmod{p}, \\ x \equiv 1 \pmod{n/p}. \end{cases}$$

By the Chinese Remainder Theorem this system has a solution b. Then by

$$\left(\frac{b}{p}\right) = -1$$

and

$$\left(\frac{b}{n/p}\right) = 1$$

we obtain

$$\left(\frac{b}{n}\right) = -1.$$

On the other hand, if

$$b^{\frac{n-1}{2}} \equiv -1 \pmod{n}$$

would hold, then certainly

$$b^{\frac{n-1}{2}} \equiv -1 \pmod{n/p}$$

would also be valid, which in view of

$$1 \equiv b \equiv b^{\frac{n-1}{2}} \pmod{n/p}$$

is not possible. Thus n is not an Euler-pseudoprime with respect to b.

So we proved that in any case, one can always find an integer b coprime to n, such that n is not a pseudoprime with respect to b. Let now b be such an integer, and let

$$1 = b_1 < \ldots < b_s \leq n-1$$

be those bases, for which n is an Euler-pseudoprime. Consider the integers

$$b \cdot b_i \quad (i = 1, \ldots, s).$$

On the one hand, they belong to different residue classes modulo n. On the other hand, if

$$(b \cdot b_i)^{\frac{n-1}{2}} \equiv \left(\frac{b \cdot b_i}{n}\right) \pmod{n}$$

would hold, then

$$b^{\frac{n-1}{2}} b_i^{\frac{n-1}{2}} \equiv \left(\frac{b}{n}\right)\left(\frac{b_i}{n}\right) \pmod{n}$$

would yield

$$b^{\frac{n-1}{2}} \equiv \left(\frac{b}{n}\right) \pmod{n},$$

which is a contradiction. Hence we get that n is not an Euler-pseudoprime with respect to any of the bases $b \cdot b_i$ $(i = 1, \ldots, s)$. From this the statement follows. □

The above theorem implies that there are no Carmichael-numbers of Euler-type. So we get the following probability prime test.

The Solovay-Strassen prime test. Let $n > 1$ be an odd integer.

- Choose randomly an integer b from the interval $(1, n)$.
- Check whether $\gcd(b, n) = 1$ holds or not. If not, then n is composite and the algorithm terminates.

- If yes, then check whether

$$b^{\frac{n-1}{2}} \equiv \left(\frac{b}{n}\right) \pmod{n} \tag{5.1}$$

  holds or not. If not, then n is composite and the algorithm terminates.

- If (5.1) holds, then we repeat the process.

- If after k steps we still could not prove that n is composite, then we can say that n is a prime with probability at least $1 - (1/2)^k$.

Remark 5.3.2 It is well-known that to calculate

$$b^{\frac{n-1}{2}} \pmod{n}$$

we need $O(\log^3 n)$ bit operations. On the other hand, to calculate the Jacobi symbol $\left(\frac{b}{n}\right)$ we do not have to find the prime factorization of n (which would make the Solovay-Strassen prime test a complete nonsense), that also can be obtained with $O(\log^3 n)$ bit operations, by the help of quadratic reciprocity. We illustrate it through the following example:

$$\left(\frac{1872}{7411}\right) = \left(\frac{16}{7411}\right) \cdot \left(\frac{117}{7411}\right) = (-1)^{\frac{(117-1)(7411-1)}{4}} \left(\frac{7411}{117}\right) =$$

$$= \left(\frac{7411}{117}\right) = \left(\frac{40}{117}\right) = \left(\frac{8}{117}\right) \cdot \left(\frac{5}{117}\right) = (-1)^{\frac{(117^2-1)}{8}} \left(\frac{5}{117}\right) =$$

$$= -\left(\frac{5}{117}\right) = (-1)^{\frac{(5-1)(117-1)}{4}} \left(\frac{117}{5}\right) = -\left(\frac{117}{5}\right) = -\left(\frac{2}{5}\right) = 1.$$

Altogether, the Solovay-Strassen prime test requires $O(\log^3 n)$ bit operations.

5.4 Strong pseudoprimes and the Miller-Rabin prime test

In this section, by a further refinement of pseudoprimality, we obtain a more efficient prime test.

Definition 5.4.1 *Let n be an odd composite number, and let $n-1 = 2^s t$, where $s > 0$, t is odd. If b is an integer with $\gcd(b, n) = 1$ such that either*

$$b^t \equiv 1 \pmod{n},$$

or

$$b^{2^r t} \equiv -1 \pmod{n}$$

for some $0 \leq r < s$, then we say that n is a strong pseudoprime with respect to b.

Remark 5.4.1 If

$$b^{2^r t} \equiv -1 \pmod{n}$$

for some $0 \leq r < s$, then

$$b^{2^i t} \equiv 1 \pmod{n}$$

for every $r < i < s$.

Our first theorem is rather important: it shows that the primes satisfy the conditions in the above definition.

Theorem 5.4.1 *Let p be an odd prime. Then p satisfies the conditions in the definition above, for arbitrary $b \in \mathbb{Z}$ with $p \nmid b$.*

Proof. Let b be an arbitrary integer not divisible with p. Then by the Euler-Fermat theorem

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

That is, if $p-1 = 2^s t$ where $s > 0$, t is odd, then with this notation we have

$$b^{2^s t} \equiv 1 \pmod{p}.$$

This implies that

$$p \mid \left(b^{2^{s-1} t} + 1\right)\left(b^{2^{s-1} t} - 1\right),$$

so

$$b^{2^{s-1} t} \equiv -1 \pmod{p} \quad \text{or} \quad b^{2^{s-1} t} \equiv 1 \pmod{p}$$

holds. In the first case we are already done. In the second case we have two options. If $s = 1$, then we are done again. On the other hand, if $s > 1$, then repeating the previous argument we obtain that

$$p \mid \left(b^{2^{s-2}t} + 1\right)\left(b^{2^{s-2}t} - 1\right),$$

whence

$$b^{2^{s-2}t} \equiv -1 \pmod{p} \quad \text{or} \quad b^{2^{s-2}t} \equiv 1 \pmod{p}.$$

Continuing this procedure, ultimately our statement follows. □

Our next theorems show that the strong pseudoprime property is indeed 'stronger' then the pseudoprime properties introduced earlier. The first statement of this flavor concerns the original pseudoprimality.

Theorem 5.4.2 *If n is a strong pseudoprime with respect to b, then n is a pseudoprime with respect to b.*

Proof. If

$$b^t \equiv 1 \pmod{n}$$

then

$$b^{\frac{n-1}{2}} \equiv b^{2^s t} \equiv 1 \pmod{n}.$$

On the other hand, if

$$b^{2^r t} \equiv -1 \pmod{n}$$

with $r < s$, then

$$b^{n-1} = b^{2^{s-r}2^r t} \equiv 1 \pmod{n}$$

follows. □

Now we show that the strong pseudoprime property also implies the Euler pseudoprime property.

Theorem 5.4.3 *If n is a strong pseudoprime with respect to b, then n is an Euler pseudoprime with respect to b.*

Proof. Let n be a strong pseudoprime with respect to b. We have to show that then

$$b^{\frac{n-1}{2}} \equiv \left(\frac{b}{n}\right) \pmod{n}. \tag{5.2}$$

Put again $n-1=2^s t$, with $s>0$, t odd. We distinguish several cases.

i) Assume first that

$$b^t \equiv 1 \pmod{n}.$$

Then certainly the left hand side of (5.2) is 1, as well. At the same time,

$$1=\left(\frac{1}{n}\right)=\left(\frac{b^t}{n}\right)=\left(\frac{b}{n}\right)^t=\left(\frac{b}{n}\right)$$

is also valid, so in this case our statement holds.

ii) Let now

$$b^{\frac{n-1}{2}} \equiv -1 \pmod{n}.$$

Let p be a prime divisor of n, and $p-1=2^u v$ with $u>0$ and v odd. We prove that then

$$u \geq s \quad \text{and} \quad \left(\frac{b}{p}\right)=\begin{cases}-1, & \text{if } u=s,\\ 1, & \text{if } u>s.\end{cases} \tag{5.3}$$

For this, we start from the assertion

$$b^{\frac{n-1}{2}} \equiv b^{2^{s-1}t} \equiv -1 \pmod{n}.$$

From this, by raising to the v-th power we obtain

$$\left(b^{2^{s-1}v}\right)^t \equiv \left(b^{2^{s-1}t}\right)^v \equiv (-1)^v \equiv -1 \pmod{n}.$$

As $p \mid n$, the above congruence modulo p is also valid, that is

$$\left(b^{2^{s-1}v}\right)^t \equiv -1 \pmod{p} \tag{5.4}$$

also holds. From this we see that $u<s$ is not possible, as otherwise by the Euler-Fermat theorem we would get

$$\left(b^{2^{s-1}v}\right)^t \equiv \left(b^{2^u v}\right)^{2^{s-u-1}t} \equiv \left(b^{p-1}\right)^{2^{s-u-1}t} \equiv 1 \pmod{p},$$

yielding a contradiction. Hence we must have $u \geq s$. If $u=s$, then by the congruence (5.4) and the assertion

$$\left(\frac{b}{p}\right) \equiv b^{\frac{p-1}{2}} \equiv b^{2^{u-1}v} \pmod{p}$$

we obtain that $\left(\frac{b}{p}\right)$ cannot be 1, whence

$$\left(\frac{b}{p}\right) = -1.$$

If $u > s$, then raising the congruence (5.4) on the 2^{u-s}-th power we get that

$$\left(b^{2^{u-1}v}\right)^t \equiv 1 \pmod{p}.$$

Thus now $\left(\frac{b}{p}\right)$ cannot equal -1, so necessarily

$$\left(\frac{b}{p}\right) = 1.$$

Thus (5.3) holds indeed. Let now

$$n = p_1^{\alpha_1} \dots p_\ell^{\alpha_\ell}$$

be the prime factorization of n, and let

$$k = \sum_{u_i = s} \alpha_i,$$

where $p_i - 1 = 2^{u_i} v_i$, $u_i > 0$, v_i is odd $(i = 1, \dots, \ell)$. By the help of assertion (5.3) we obtain that

$$u_i \geq s \quad (i = 1, \dots, \ell),$$

and also

$$\left(\frac{b}{n}\right) = \prod_{i=1}^{\ell} \left(\frac{b}{p_i}\right)^{\alpha_i} = (-1)^k.$$

Observe further that

$$p_i \equiv t_i \pmod{2^{s+1}},$$

where

$$t_i = \begin{cases} 2^s + 1, & \text{if } u_i = s, \\ 1, & \text{otherwise} \end{cases} \quad (i = 1, \dots, \ell).$$

Since

$$n \equiv 1 + 2^s t \equiv 1 + 2^s \pmod{2^{s+1}},$$

we get

$$1 + 2^s \equiv \prod_{i=1}^{\ell} p_i^{\alpha_i} \equiv (1 + 2^s)^k \equiv 1 + k2^s \pmod{2^{s+1}}.$$

Therefore k is odd, so

$$\left(\frac{b}{n}\right) = (-1)^k = -1.$$

Thus our statement holds also in this case.

iii) Finally, suppose that

$$b^{2^r t} \equiv -1 \pmod{n}$$

holds with some $0 \le r < s - 1$. Then of course

$$b^{\frac{n-1}{2}} \equiv 1 \pmod{n},$$

so we need to show that

$$\left(\frac{b}{n}\right) = 1.$$

Let again p be a prime divisor of n, and $p - 1 = 2^u v$, $u > 0$, v is odd. Then

$$u \ge r + 1 \quad \text{and} \quad \left(\frac{b}{p}\right) = \begin{cases} -1, & \text{if } u = r + 1, \\ 1, & \text{if } u > r + 1 \end{cases} \tag{5.5}$$

hold. As the proof of (5.5) is very much similar to that of (5.3), it is left to the Reader. Let again

$$n = p_1^{\alpha_1} \dots p_\ell^{\alpha_\ell}$$

be the prime factorization of n, and put

$$k = \sum_{u_i = r+1} \alpha_i,$$

where $p_i - 1 = 2^{u_i} v_i$, $u_i > 0$, v_i is odd ($i = 1, \dots, \ell$). Similarly to case ii), we obtain that

$$\left(\frac{b}{n}\right) = \prod_{i=1}^{\ell} \left(\frac{b}{p_i}\right)^{\alpha_i} = (-1)^k.$$

As

$$n \equiv 1 + 2^s t \equiv 1 \pmod{2^{r+2}},$$

we get

$$1 \equiv \prod_{i=1}^{\ell} p_i^{\alpha_i} \equiv (1+2^{r+1})^k \equiv 1 + k2^{r+1} \pmod{2^{r+2}}.$$

So now k is even, that is

$$\left(\frac{b}{n}\right) = (-1)^k = 1.$$

Thus our claim follows also in this case. Hence the proof of the theorem is complete. □

Our next theorems concern special cases. First we investigate strong pseudoprimes with respect to the base 2.

Theorem 5.4.4 *There are infinitely many strong pseudoprimes with respect to the base* 2*.*

Proof. We show that if n is an (odd) pseudoprime with respect to 2, then $2^n - 1$ is a strong pseudoprime with respect to 2. As we know that there exist infinitely many pseudoprimes with respect to 2, hence this will imply the statement.

Let n be an (odd) pseudoprime with respect to 2, and let $m = 2^n - 1$. Now by

$$2^{n-1} \equiv 1 \pmod{n}$$

we obtain

$$2^{n-1} - 1 = k \cdot n,$$

where k is odd. Thus

$$m - 1 = 2^n - 2 = 2kn,$$

where kn is odd. Using that

$$m = 2^n - 1 \mid 2^{kn} - 1$$

we get

$$2^{\frac{m-1}{2}} \equiv 2^{kn} \equiv 1 \pmod{m}.$$

Since $(m-1)/2$ is odd, our statement follows. □

We give the next theorem for the sake of curiosity.

Theorem 5.4.5 *The Fermat-number $F_n = 2^{2^n}+1$ $(n \geq 0)$ is either a prime, or a strong pseudoprime with respect to 2.*

Proof. Observe that for any $n \geq 0$ we have

$$2^{F_n-1} = 2^{2^{2^n}} = \left(2^{2^n}\right)^{2^{2^n-n}} = (F_n-1)^{2^{2^n-n}}.$$

Thus

$$2^{\frac{F_n-1}{2^i}} \equiv 1 \pmod{F_n}$$

holds for $i = 0, 1, \ldots, 2^n - n - 1$, and

$$2^{\frac{F_n}{2^i}} \equiv -1 \pmod{F_n}$$

holds for $i = 2^n - n$. This proves the statement. □

The last theorem of the section - which we give without a proof - is of great importance: it shows that the strong pseudoprime property is not only 'theoretically' stronger than the Euler pseudoprime property.

Theorem 5.4.6 *Let n be an odd composite number. Then n is a strong pseudoprime with respect to at least 75% of the integers coprime n, counted modulo n.*

The Miller-Rabin prime test. Let $n > 1$ be an odd integer, and write $n - 1 = 2^s t$ with t odd.

- Choose an integer b randomly from the interval $(1, n)$.
- Check if $\gcd(b, n) = 1$ holds or not. If not, then n is composite and the algorithm terminates.
- Check whether
$$b^t \equiv \pm 1 \pmod{n}$$
holds or not. If yes, then we repeat the process with choosing another b.

- If not, then consider the number $b^{2t} \pmod{n}$. If it is -1, then we choose a new b, otherwise we consider b^{4t}, and so on. If any of the congruences

$$b^{2^r t} \equiv -1 \pmod{n} \quad (r = 1, \ldots, s-1)$$

holds, then we choose another b. On the other hand, if $b^{2^r t} \not\equiv -1 \pmod{n}$ for any of these values of r, then n is composite.

- If after k iterations we still cannot conclude that n is composite, then we can say that n is a prime with probability at least $1 - (1/4)^k$.

Remark 5.4.2 One can check that the Miller-Rabin test requires $O(\log^3 n)$ bit operations, similarly to the Solovay-Strassen prime test.

5.5 Deterministic prime tests

The first statement of the section is Wilson's theorem, which from the theoretical point of view can be considered to be a deterministic prime test.

Theorem 5.5.1 *Let $n > 1$. Then n is a prime if and only if*

$$(n-1)! \equiv -1 \pmod{n}.$$

Proof. Assume first that n is composite, however, still

$$(n-1)! \equiv -1 \pmod{n}$$

holds. Let p be a prime divisor of n; then we clearly have $2 \leq p \leq n-1$. Thus on the one hand $p \mid n$ implies

$$(n-1)! \equiv -1 \pmod{p},$$

and on the other hand by $p \mid (n-1)!$ we have

$$(n-1)! \equiv 0 \pmod{p},$$

which is a contradiction.

Let now n be a prime. As the statement trivially holds for $n = 2$, without loss of generality we may assume that $n \geq 3$. Then for any $a \in \{1, \ldots, n-1\}$

$$a^{n-1} \equiv 1 \pmod{n}$$

holds. Thus the roots of the polynomial $x^{n-1} - 1$ modulo n are precisely the numbers $1, \ldots, n-1$. However, then we have

$$x^{n-1} - 1 \equiv (x-1)\ldots(x-(n-1)) \pmod{n}.$$

Comparing the constant terms of the above polynomials, using that n is odd, we obtain

$$(n-1)! \equiv -1,$$

which was to be proved. □

Remark 5.5.1 As the calculation of $n!$ requires $O(n^2 \log^2 n)$ bit operations, the above theorem cannot be used for prime testing in practice.

Remark 5.5.2 Starting from the 1980's, Adleman, Pomerance, Rumley and Cohen, Lenstra worked out a deterministic version of the Miller-Rabin test, based upon deep algebraic number theoretical tools. The test requires $O((\log n)^{c \log\log\log n})$ bit operations, where c is an absolute constants. Agrawal, Kayal and Saxena in 2004 constructed a polynomial deterministic algorithm, which for any input n decides whether n is a prime or not.

5.6 Exercises

Exercise 5.6.1 Find all integers b, with respect to which

- $n = 15$,
- $n = 21$,
- $n = 91$

is pseudoprime!

Exercise 5.6.2 Let p be a prime, such that $2p-1$ is also prime. Show that then $n = p(2p-1)$ is a pseudoprime with respect to an integer b coprime n, provided that

$$\left(\frac{b}{2p-1}\right) = 1$$

holds! Using this assertion, prove that n is a pseudoprime precisely for half of the bases coprime to n, counted modulo n.

Exercise 5.6.3 Let p be a prime. Show that p^2 is pseudoprime with respect to an integer b if and only if

$$b^{p-1} \equiv 1 \pmod{p^2}$$

holds!

Exercise 5.6.4 Let p, q be primes, $d = \gcd(p-1, q-1)$, and put $n = pq$. Prove that n is pseudoprime with respect to an integer b if and only if

$$b^d \equiv 1 \pmod{n}$$

holds!

Exercise 5.6.5 Let $m \in \mathbb{N}$ such that $6m+1$, $12m+1$, $18m+1$ are all primes. Prove that then

$$n = (6m+1)(12m+1)(18m+1)$$

is a Carmichael-number!

Exercise 5.6.6 Verify that the following numbers are Carmichael-numbers:

- $1105 = 5 \cdot 13 \cdot 17$
- $1729 = 7 \cdot 13 \cdot 19$
- $2465 = 5 \cdot 17 \cdot 29$
- $2821 = 7 \cdot 13 \cdot 31$
- $6601 = 7 \cdot 23 \cdot 41$
- $29341 = 13 \cdot 37 \cdot 61$

- $172081 = 7 \cdot 13 \cdot 31 \cdot 61$
- $278545 = 5 \cdot 17 \cdot 29 \cdot 113$

Exercise 5.6.7 Show that 561 is the smallest Carmichael-number!

Exercise 5.6.8 Let p, q be primes. Find all Carmichael-numbers of the form $3pq$.

Exercise 5.6.9 Let p, q be primes. Find all Carmichael-numbers of the form $5pq$.

Exercise 5.6.10 Show that for any fixed prime r there exist only finitely many primes p, q such that rpq is a Carmichael-number!

Exercise 5.6.11 Find all integers b with respect to which $n = 561$ is an Euler pseudoprime!

Exercise 5.6.12 Find all integers b with respect to which $n = 561$ is a strong pseudoprime!

Exercise 5.6.13 Show that 65 is a strong pseudoprime with respect to both 8 and 18, but it is not a strong pseudoprime with respect to $144 = 8 \cdot 18$.

Exercise 5.6.14 Let p be a prime, $n = p^k$ with $k > 1$. Prove that then n is a strong pseudoprime with respect to an integer b if and only if n is a pseudoprime with respect to b.

Exercise 5.6.15 Let p be an odd prime and a be an integer such that

$$a \not\equiv \pm 1 \pmod{p}.$$

Prove that then the number

$$n = \frac{a^{2p} - 1}{a^2 - 1}$$

is pseudoprime with respect to a.

Exercise 5.6.16 Let $k \geq 1$, r odd with $0 < r < 2^k$, and let

$$n = 2^k r + 1.$$

Prove that if for some integer a we have

$$a^{\frac{n-1}{2}} \equiv -1 \pmod{n},$$

then n is a prime!

Exercise 5.6.17 Code the Solovay-Strassen prime test! Run the test for all n in the interval $[10000, 20000]$, using $k = 1, 2, 5, 10$ iterations! If we write the code in the program package Maple (or in some other mathematical program package), then using the internal functions determine that how many composite numbers passed the tests!

Exercise 5.6.18 Code the Miller-Rabin prime test! Run the test for all n in the interval $[10000, 20000]$, using $k = 1, 2, 5, 10$ iterations! If we write the code in the program package Maple (or in some other mathematical program package), then using the internal functions determine that how many composite numbers passed the tests! Compare the results with those of the previous exercise.

Exercise 5.6.19 What is the remainder of the product

$$8 \cdot 9 \cdot 10 \cdot 11 \cdot 12 \cdot 13$$

after division by 7?

Exercise 5.6.20 Prove that for any prime p

$$p^2 \mid ((2p-1)! - p)$$

holds!

Exercise 5.6.21 Let p be a prime. What is the remainder of the number

$$(p+1)! - p!$$

after division by p^3 ?

Exercise 5.6.22 Prove that for any prime p

$$p^p \mid ((p^2-1)! - p^{p-1})$$

holds!

Exercise 5.6.23 Prove that for any odd prime p

$$\left(\left(\frac{p-1}{2}\right)!\right)^2 \equiv (-1)^{\frac{p-1}{2}} \pmod{p}$$

holds!

Exercise 5.6.24 Prove that for any odd prime p and $a \in \mathbb{Z}$

$$p \mid a^p + a \cdot (p-1)!$$

holds!

Exercise 5.6.25 Prove that for infinitely many primes p one can find an $n \in \mathbb{N}$ for which

$$p \mid ((n-1)! + 1)$$

holds!

Chapter 6

Factorization algorithms

In this chapter we discuss methods and procedures providing the prime factorization of a given positive integer n. In fact, formally we shall do less: we shall be content with finding a non-trivial divisor d of n. It is sufficient: if we succeed, then repeating the procedure with n/d in place of n, finally we get the prime factorization of n.

We also mention that in many cases we shall assume that n is odd and composite. Obviously, the parity of n can be trivially checked. On the other hand, based upon the previous chapter, we can decide efficiently (with a polynomial algorithm) whether n is a prime or not. So these assumptions can be made without loss of generality.

6.1 Trial division

First we outline a classical (though rather primitive) procedure, the so-called trial division.

Let n be composite. Divide n successively by the numbers $2, 3, \ldots, \lfloor\sqrt{n}\rfloor$ until we get an integer quotient. Since n must have a divisor d with $2 \leq d \leq \lfloor\sqrt{n}\rfloor$, thus the procedure leads to a factorization of n. This is a nice and simple, but not efficient algorithm, of complexity $O(\sqrt{n})$.

6.2 Pollard's ϱ-method

Before outlining the method we describe an interesting phenomenon, the so-called birthday paradox, which is a kind of background of the procedure.

The birthday paradox. At least how many people need to be in a room for that it is better to bet on that

i) at least one of them was born on a particular day, say on 1 January?

ii) there are two of them having their birthdays on the same day?

Write k for the number of people in the room. For simplicity, assume that nobody was born on 29 February. (It is not an important point, only makes the calculations simpler.)

In part i) we are looking for the smallest value of k_1 with

$$1 - \left(\frac{364}{365}\right)^{k_1} \geq \frac{1}{2}.$$

A simple calculation yields $k_1 = 253$.

In part ii) we want to find the smallest value of k_2 with

$$1 - \frac{365(365-1)\dots(365-(k_2-1))}{365^{k_2}} \geq \frac{1}{2}.$$

Hence we easily get $k_2 = 23$.

The two values are surprisingly far away, in particular, the value $k_2 = 23$ may seem to be extremely low. The paradox shows that it can be much more easy to find two elements which are similar from some aspect, then to find one exactly holding some specific property.

Pollard's ϱ-method. Let be given a composite number $n > 1$, which we want to factorize. With the birthday paradox in mind, we are not looking for an integer d with $d \mid n$, or in other words, an integer t with

$$t \equiv 0 \pmod{d}$$

where $d \mid n$, rather we intend to find integers t_1, t_2 such that

$$t_1 \equiv t_2 \pmod{d},$$

where $d \mid n$. Then clearly $\gcd(t_1 - t_2, n)$ is a divisor of n.

To find integers t_1, t_2 with the above property, do the following. Take a polynomial $f(x)$ of degree at least two, and an integer x_0. Consider the following sequence:

$$x_1 = f(x_0) \pmod{n}, x_2 = f(x_1) \pmod{n}, \dots, x_k = f(x_{k-1}) \pmod{n}, \dots,$$

where the polynomial values obtained are reduced modulo n in every step. After calculating x_i ($i \geq 1$), check whether

$$n > \gcd(x_i - x_{i-1}, n) > 1$$

holds. If yes, then we clearly get a factorization of n. If not, then we continue the process with step $i + 1$.

Example 6.2.1 Let $n = 91$. Take the polynomial $f(x) = x^2 + 1$ and the integer $x_0 = 1$. Then

$$x_1 = 2, \quad x_2 = 5, \quad x_3 = 26.$$

Hence we get

$$\gcd(2 - 1, 91) = 1, \quad \gcd(5 - 2, 91) = 1, \quad \gcd(26 - 5, 91) = 7$$

whence we obtain the factorization $91 = 7 \cdot 13$.

Remark 6.2.1 We make several notes.

1. The average bit operation requirement of the process is $O(\sqrt[4]{n} \log^3 n)$.
2. The method has several variants. For example, in step i we can calculate all the values

   $$\gcd(x_i - x_{i-1}, n), \ \gcd(x_i - x_{i-2}, n), \ \ldots, \gcd(x_i - x_0, n),$$

   hoping that one of these values falls (sharply) between 1 and n.
3. It is important that the degree of the polynomial f taken is at least two. Indeed, if f is linear, then the iterated values $f(x_i)$ do not give a sufficient 'shuffle' of the residue classes modulo n.

6.3 Fermat-factorization

In this section we discuss a factorization technique going back to Fermat. The process is based upon the following theorem, establishing a bijective mapping between the representations of an odd integer as a product of two terms, and as a difference of two squares.

Theorem 6.3.1 *Let n be an odd positive integer. Then there is a bijective mapping between the representations*

$$n = ab \quad (a \geq b > 0)$$

and the representations

$$n = t^2 - s^2 \quad (t > s \geq 0).$$

One such bijection is given by the following formulas:

$$t = \frac{a+b}{2}, \quad s = \frac{a-b}{2}$$

and

$$a = t + s, \quad b = t - s.$$

Proof. First let n be of the form

$$n = ab,$$

where $a \geq b > 0$. Then clearly, both a and b are odd, thus

$$n = \left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2$$

is a representation of n as the difference of two squares. In other words, we can write

$$n = t^2 - s^2,$$

where

$$t = \frac{a+b}{2}, \quad s = \frac{a-b}{2},$$

and hence $t \geq s \geq 0$.

On the other hand, if

$$n = t^2 - s^2$$

holds with $t > s \geq 0$, then we clearly have

$$n = (t+s)(t-s).$$

So by the notation

$$a = t + s, \quad b = t - s,$$

we can write

$$n = ab$$

with $a \geq b > 0$. This proves the statement. $\square$

Fermat-factorization. Let $n > 1$ be an odd composite integer, which we want to factorize. By the above assertion, instead of trying to represent n as a product, we find a representation of n as the difference of two squares. If n is of the form

$$n = ab$$

such that a and b are 'close' to each other, then in the representation

$$n = t^2 - s^2$$

the integer

$$s = \frac{a-b}{2}$$

is 'small', and

$$t = \frac{a+b}{2}$$

is just 'slightly' larger than $\sqrt{n}$. Therefore checking the numbers

$$t = \lceil \sqrt{n} \rceil,\ \lceil \sqrt{n} \rceil + 1,\ \lceil \sqrt{n} \rceil + 2,\ \ldots$$

after a while we get that

$$t^2 - n = s^2$$

is a square, whence we immediately obtain a factorization of n. The algorithm (after certain refinements not outlined here) requires $O(\sqrt[3]{n})$ bit operations.

Example 6.3.1 Factorize the integer $n = 200819$. We obtain the following:

$$\lceil \sqrt{n} \rceil = 449, \quad 449^2 - 200819 = 782 \text{ is not a square;}$$

$$\lceil \sqrt{n} \rceil + 1 = 450, \quad 450^2 - 200819 = 1681 = 41^2.$$

Thus

$$n = 450^2 - 41^2 = 491 \cdot 409.$$

6.4 Factorbase factorization

Observe that the heart of the Fermat-factorization is the following. To factorize n, we have to find integers t, s such that

$$t^2 \equiv s^2 \pmod{n},$$

however,

$$t \not\equiv \pm s \pmod{n}.$$

Indeed, then we have

$$n \mid (t+s)(t-s),$$

but

$$n \nmid t+s, \quad n \nmid t-s,$$

and so calculating $\gcd(n, t+s)$ and $\gcd(n, t-s)$ (by the help of the Euclidean algorithm) we get a factorization of n.

Example 6.4.1 Let $n = 4633$. We observe that

$$118^2 \equiv 25 \equiv 5^2 \pmod{n}.$$

Hence by

$$\gcd(4633, 118+5) = 41$$

and

$$\gcd(4633, 118-5) = 113$$

we obtain that

$$4633 = 41 \cdot 113$$

holds.

To be able to apply the above principle, we have to answer the following two questions.

1. Do there always exist integers t and s holding the above properties?
2. If yes, then how can we find such numbers (efficiently)?

The next theorem gives a positive answer to our first question.

Theorem 6.4.1 *Let $n > 1$ be odd, and let*

$$n = p_1^{\alpha_1} \dots p_r^{\alpha_r}$$

be the prime factorization of n. Then for any integer a coprime to n the congruence

$$x^2 \equiv a^2 \pmod{n}$$

has precisely 2^r solutions modulo n.

Proof. Take first $r = 1$: this particular case needs a special treatment. Then $n = p_1^{\alpha_1}$, and

$$p_1^{\alpha_1} \mid x^2 - a^2$$

implies

$$p_1^{\alpha_1} \mid (x + a)(x - a).$$

Then (as n is odd and $\gcd(n, a) = 1$) we have

$$p_1 \nmid 2a.$$

Hence

$$p_1^{\alpha_1} \mid x + a \quad \text{or} \quad p_1^{\alpha_1} \mid x - a,$$

whence

$$x \equiv \pm a \pmod{n},$$

which proves our claim in this case.

Let now $r \geq 2$. For every $i = 1, \dots, r$ take $\varepsilon \in \{-1, 1\}$. Obviously, the number of the tuples

$$(\varepsilon_1, \dots, \varepsilon_r)$$

obtained in this way is just 2^r. Consider the linear congruence system

$$\begin{cases} x & \equiv \varepsilon_1 \pmod{p_1^{\alpha_1}} \\ & \vdots \\ x & \equiv \varepsilon_r \pmod{p_r^{\alpha_r}} \end{cases}.$$

Observe that its solutions coincide with the solutions of the congruence

$$x^2 \equiv a^2 \pmod{n}. \tag{6.1}$$

Indeed, if x is a solution of the above congruence system, then by

$$x^2 \equiv a^2 \pmod{p_i^{\alpha_i}} \quad (i = 1, \ldots, r)$$

it is also a solution to the congruence (6.1). On the other hand, if x is a solution of (6.1), then by what we have proved in case of $r = 1$, we have

$$x \equiv \pm a \pmod{p_i^{\alpha_i}} \quad (i = 1, \ldots, r).$$

That is, x is also a solution of the above congruence system. By the Chinese Remainder Theorem the congruence system has precisely on solution modulo n, and it is also obvious that the congruence systems corresponding to different choices of the ε_i have different solutions. As the number of congruence systems in question is 2^r, this implies the statement. □

Factorbase factorization. Now we are ready to answer our second question above. Before that, we introduced some more notation.

Let $n > 1$ be an odd integer. In the rest of this chapter for $a \in \mathbb{Z}$ we write $a \pmod{n}$ for that uniquely determined integer b, for which

$$a \equiv b \pmod{n} \quad \text{and} \quad -\frac{n}{2} < b < \frac{n}{2}$$

hold. This integer b is usually called the *absolute residue* of a.

Let

$$B = \{p_1, \ldots, p_k\} \cup \{-1\},$$

where $p_1, \ldots, p_k$ are distinct primes. The set B is called a factorbase. An integer a is called a B-integer, if all the prime divisors of $a \pmod{n}$ belong to B.

By the above notation, factorbase factorization works in the following way. Let n be an odd composite integer which is not a prime power. (Note that one can obviously check in polynomial time whether n is a prime power or not.) Choose a 'medium' size integer y (e.g., if n has 50 digits then let y have 5-6 digits), and set

$$B = \{p \text{ is a prime} : p \leq y\} \cup \{-1\}$$

be our factorbase. Choose randomly a certain number of integers b_i, and decide whether $b_i^2 \pmod{n}$ is a B-integer or not. If yes, then note the factorization

$$b_i^2 \pmod{n} = (-1)^{\alpha_{i0}} p_1^{\alpha_{i1}} \ldots p_k^{\alpha_{ik}},$$

where $\alpha_{i0} \in \{0, 1\}$, and the exponents $\alpha_{i1}, \ldots, \alpha_{ik}$ are non-negative integers. Further, let

$$v_i = (\alpha_{i0} \pmod 2, \alpha_{i1} \pmod 2, \ldots, \alpha_{ik} \pmod 2),$$

that is, the entries of v_i are zeroes and ones according to the parities of the above exponents. Find a linear combination of the vectors v_i with coefficients $0, 1$, which yields the zero vector modulo 2. (This can be efficiently done by linear algebraic tools over the field of two element.) If we succeed, then with some finite index set I we get that each component of the vector

$$\sum_{i \in I} v_i$$

is even. Thus

$$(-1)^{\sum_{i\in I} \alpha_{i0}} p_1^{\sum_{i\in I} \alpha_{i1}} \ldots p_k^{\sum_{i\in I} \alpha_{ik}} = \prod_{i \in I} (b_i^2 \pmod n) \equiv \prod_{i \in I} b_i^2 \pmod n$$

holds - with squares on both sides! That is, we obtained a congruence of the form

$$c^2 \equiv d^2 \pmod n.$$

If here we have

$$c \not\equiv \pm d \pmod n,$$

then based upon the above considerations we get a factorization of n. As we obtained c and d 'independently', we have a good chance for that. If we are unlucky and here $c \equiv \pm d \pmod n$ holds, then we continue the procedure by choosing new values b_i. The process requires $O(e^{C\sqrt{\log n \log\log n}})$ bit operations, where C is a value depending on the parameters.

Example 6.4.2 Let $n = 1829$ and

$$B = \{-1, 2, 3, 5, 7, 11, 13\}.$$

After some calculations we obtain the following B-integers and factorizations:

$$b_1^2 = 42^2 \equiv (-1) \cdot 5 \cdot 13 \pmod{1829},$$

$$b_2^2 = 43^2 \equiv 2^2 \cdot 5 \pmod{1829},$$

$$b_3^2 = 61^2 \equiv 3^2 \cdot 7 \pmod{1829},$$

$$b_4^2 = 74^2 \equiv (-1) \cdot 11 \pmod{1829},$$
$$b_5^2 = 85^2 \equiv (-1) \cdot 7 \cdot 13 \pmod{1829},$$
$$b_6^2 = 86^2 \equiv 2^4 \cdot 5 \pmod{1829}.$$

From this the exponent vectors v_i modulo 2 are:

$$v_1 = (1,0,0,1,0,0,1), \quad v_2 = (0,0,0,1,0,0,0),$$
$$v_3 = (0,0,0,0,1,0,0), \quad v_4 = (1,0,0,0,0,1,0),$$
$$v_5 = (1,0,0,0,1,0,1), \quad v_6 = (0,0,0,1,0,0,0).$$

We may immediately observe that

$$v_2 + v_6 \equiv (0,0,0,0,0,0,0) \pmod{2},$$

where the congruence is to be understood componentwise. Thus we obtain the assertion

$$2^6 \cdot 5^2 = 40^2 \equiv (43 \cdot 86)^2 \pmod{1829}.$$

But we are unlucky:

$$40 \equiv 43 \cdot 86 \pmod{1829},$$

so the congruence is trivial, it does not lead to a factorization of 1829. However, we also have the assertion

$$v_1 + v_2 + v_3 + v_5 \equiv (0,0,0,0,0,0,0) \pmod{2},$$

whence

$$901^2 \equiv (-1)^2 \cdot 2^2 \cdot 3^2 \cdot 5^2 \cdot 7^2 \cdot 13^2 \equiv$$
$$\equiv (42 \cdot 43 \cdot 61 \cdot 85)^2 \equiv 1459^2 \pmod{1829}$$

follows. Since here

$$901 \not\equiv 1459 \pmod{1829},$$

thus

$$\gcd(1459 + 901, 1829) = 59$$

and

$$\gcd(1459 - 901, 1829) = 31$$

yield the following factorization of 1829:

$$1829 = 31 \cdot 59.$$

6.5 Exercises

Exercise 6.5.1 Factorize n with Pollard's ϱ-method, using the given polynomial $f(x)$ and integer x_0:

a) $f(x) = x^2 - 1$, $x_0 = 2$, $n = 91$,

b) $f(x) = x^2 + 1$, $x_0 = 1$, $n = 8051$,

c) $f(x) = x^2 - 1$, $x_0 = 5$, $n = 7031$,

d) $f(x) = x^3 + x + 1$, $x_0 = 1$, $n = 2701$.

Exercise 6.5.2 Factorize the following integers with Fermat-factorization:

a) 8633

b) 809009

c) 92296873

d) 88169891

e) 4601

Exercise 6.5.3 Let $n = 2701$. Choose an appropriate factorbase B, for which the numbers 52^2 and 53^2 are B-integers, and using this factorize n.

Chapter 7

Fermat's equation and Pythagorean triples

In this chapter we study the solutions of the equation

$$x^n + y^n = z^n \tag{7.1}$$

in positive integers x, y, z, where $n \geq 2$. As we shall see, the cases $n = 2$ and $n \geq 3$ are fundamentally different.

7.1 Fermat's equation

Consider the equation (7.1). Fermat in 1637 formulated his famous conjecture stating that for $n \geq 3$ the equation has no solutions in positive integers. The case $n = 4$ was proved by Fermat himself (with the famous method of 'infinite descent'), while the proof of the conjecture for $n = 3$ is due to Euler. The conjecture in its full generality was verified by Wiles in 1995.

7.2 Pythagorean triples

Consider now the equation

$$x^2 + y^2 = z^2 \tag{7.2}$$

in positive integers x, y, z. Finding the solutions of the equation was already done by the ancient Greeks. Observe that if x, y, z is a solution, then by the

theorem of Pythagoras, x, y, z are the length of the sides of a right triangle having integral sides. We introduce the following notion.

Definition 7.2.1 *The positive integer solutions x, y, z of equation (7.2) are called Pythagorean triples. If we have $\gcd(x, y, z) = 1$, then we say that the triple is reduced.*

Remark 7.2.1 We mention that having the Pythagorean triples, all integer solutions of (7.2) can be easily described.

Further, if x, y, z is a reduced Pythagorean triple, then for any positive integer d the numbers dx, dy, dz form a Pythagorean triple. On the other hand, the reverse statement is also valid, since

$$\gcd(x, y, z) = d$$

implies that

$$\frac{x}{d}, \ \frac{y}{d}, \ \frac{z}{d}$$

is a reduced Pythagorean triple. Thus to describe all Pythagorean triples, it is sufficient to characterize all reduced Pythagorean triples.

Our next theorem provides a precise description (so-called parametrization) of the reduced Pythagorean triples.

Theorem 7.2.1 *Let x, y, z be a reduced Pythagorean triple. Then we have*

$$x = u^2 - v^2, \quad y = 2uv, \quad z = u^2 + v^2$$

with some positive integers u, v, for which

$$u > v, \quad \gcd(u, v) = 1, \quad u \not\equiv v \pmod 2$$

hold. Here the roles of x and y can be switched.

Proof. First we show that the integers x, y, z obtained by the above parametrization, form reduced Pythagorean triples. It can be easily seen that x, y, z forms a solution of (7.2), since

$$x^2 + y^2 = (u^2 - v^2)^2 + (2uv)^2 = u^4 + 2u^2v^2 + v^4 = (u^2 + v^2)^2 = z^2.$$

Thus x, y, z is a Pythagorean triple. If this triple would not be reduced, then with some prime p by $p \mid x, z$ we would have $p \mid z + x, z - x$, whence $p \mid 2u^2, 2v^2$. However, it is not possible as on the one hand $\gcd(u, v) = 1$, and on the other hand by $u \not\equiv v \pmod 2$ we get that x and z are odd. Thus x, y, z is a reduced Pythagorean triple, indeed.

Let now x, y, z be a reduced Pythagorean triple. We show that we have the parametrization in the theorem. First we note that by

$$\gcd(x, y, z) = 1$$

we have in fact that

$$\gcd(x, y) = \gcd(y, z) = \gcd(x, z) = 1.$$

Indeed, if the above equalities would not hold, and e.g. say $\gcd(x, y) > 1$ would be valid, then x, y would have a common prime factor p. However, then by (7.2) the relation $p \mid z$ would also hold. One can similarly see that $\gcd(y, z) > 1$ and $\gcd(x, z) > 1$ are also not possible.

Observe now that precisely one of x, y, z is even. If it would be z, then (7.2) modulo 4 would yield

$$1 + 1 \equiv 0 \pmod 4,$$

which is impossible. Thus one of x, y, say y is even, and x and z are odd. (Clearly, by symmetry the case where x is even is similar.) Now rewrite (7.2) as

$$y^2 = z^2 - x^2.$$

Then by factorization and dividing by 4 we obtain that

$$\left(\frac{y}{2}\right)^2 = \frac{z + x}{2} \cdot \frac{z - x}{2}. \tag{7.3}$$

Observe that here all of $y, z + x, z - x$ are even, so the numbers occurring in the above equation are integers. Further, we also have

$$\gcd\left(\frac{z + x}{2}, \frac{z - x}{2}\right) = 1. \tag{7.4}$$

Indeed, if

$$\gcd\left(\frac{z + x}{2}, \frac{z - x}{2}\right) = d > 1$$

would hold, then by

$$d \mid \frac{z+x}{2} \quad \text{and} \quad d \mid \frac{z-x}{2}$$

we would get

$$d \mid \frac{z+x}{2} + \frac{z-x}{2} = z \quad \text{and} \quad d \mid \frac{z+x}{2} - \frac{z-x}{2} = x,$$

which is not possible. Thus (7.3) can hold if and only if both terms on the right hand side are full squares, that is

$$\frac{z+x}{2} = u^2, \quad \frac{z-x}{2} = v^2$$

hold with some positive integers u, v. Here by (7.4) we also have $\gcd(u, v) = 1$. Then taking the sum and difference of the above two expressions we obtain

$$z = u^2 + v^2 \quad \text{and} \quad x = u^2 - v^2.$$

Since $x > 0$, we also have $u > v$. Hence, using e.g. (7.3) we get

$$y = 2uv.$$

Thus we only need to show that $u \not\equiv v \pmod 2$. Assume to the contrary that this assertion is not valid. Then by $\gcd(u, v) = 1$ we obtain that both u and v are odd. However, it is impossible, since otherwise both x and z would be even - but $\gcd(x, z) = 1$. Hence the proof of the theorem is complete. □

7.3 Exercises

Exercise 7.3.1 Give a Pythagorean triple containing 198. Does there exist a reduced Pythagorean triple containing 198?

Exercise 7.3.2 Find all Pythagorean triples containing the element

- 56
- 35
- 42

Exercise 7.3.3 Prove that for any $n \in \mathbb{N}$ one can find a Pythagorean triple containing n.

Exercise 7.3.4 Describe those $n \in \mathbb{N}$ for which one can find a reduced Pythagorean triple containing n.

Exercise 7.3.5 Prove that there are no reduced Pythagorean triples consisting of primes!

Exercise 7.3.6 Prove that if x, y, z is a Pythagorean triple then $12 \mid xy$.

Exercise 7.3.7 Show that if x, y, z is a Pythagorean triple then $5 \mid xyz$.

Exercise 7.3.8 Let $n \geq 2$. Prove that the equation

$$x^n + y^n = z^n$$

has no positive integer solution x, y, z forming an arithmetic progression!

Chapter 8

Elements of lattice theory

In this chapter we discuss the basics of lattice theory. This area is interesting in itself, however - as we shall see - it also has important applications, as well.

8.1 Basic notions

Definition 8.1.1 *Let $n \geq k \geq 1$, and let $a_1, \ldots, a_k \in \mathbb{R}^n$ be linearly independent vectors over $\mathbb{R}$. Then the set*

$$\Lambda = \{x_1 a_1 + \cdots + x_k a_k : x_1, \ldots, x_k \in \mathbb{Z}\}$$

is called a lattice, and we say that the vectors $a_1, \ldots, a_k$ form a generating system or a basis of the lattice. We also use the phrase that the lattice Λ is generated by the vectors $a_1, \ldots, a_k$. If here $k = n$ holds, then we say that the lattice is full.

Example 8.1.1 Let $a_1 = (1, 0)$, $a_2 = (0, 1)$. Then

$$\Lambda = \{x_1 a_1 + x_2 a_2 : x_1, x_2 \in \mathbb{Z}\} = \mathbb{Z}^2$$

is the well-known square lattice.

Before proceeding further, we verify an important property of lattices.

Theorem 8.1.1 *Let Λ be a lattice in $\mathbb{R}^n$. Then $(\Lambda, +)$ is an Abelian group.*

Proof. The proof is trivial, so we only list the main points of the argument. By the definition of the set Λ it is obvious that Λ is closed for addition. The zero vector belongs to Λ. It is also clear that for any $a \in \Lambda$ the inverse of a (that is, $-a$) is also in Λ. Finally, all the required properties related to addition and the inverse are fulfilled - as they are valid in the larger set $\mathbb{R}^n$. $\square$

Beside the above, fundamental algebraic property latices hold a rather important analytic (metric) property, as well. This property is the following.

Definition 8.1.2 *Let $T \subseteq \mathbb{R}^n$. If T has no accumulation point belonging to T (in $\mathbb{R}^n$), then the set T is called discrete.*

Theorem 8.1.2 *Let Λ be a lattice in $\mathbb{R}^n$. Then Λ is discrete.*

Proof. As every lattice is a subset of a full lattice, we may clearly assume that Λ is a full lattice. Let $a_1, \ldots, a_n$ be a generating system of Λ, and let $u, v \in \Lambda$. Assume that

$$|u - v| < \varepsilon,$$

where ε is a positive real number to be specified later. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be the coefficients of u and v, respectively, with respect to the generating system $a_1, \ldots, a_n$; these coefficients are clearly integers. Then we can write

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \quad \text{and} \quad A \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

where A is the matrix of type $n \times n$, whose i-th column is the vector a_i $(i = 1, \ldots, n)$, and the vectors on the right hand sides are u and v (written as column vectors). Since $|u - v| < \varepsilon$, we have

$$|u_i - v_i| < \varepsilon \quad (i = 1, \ldots, n).$$

As $a_1, \ldots, a_n$ is a basis of $\mathbb{R}^n$, so the matrix A is an invertible matrix; let

$$A^{-1} = \begin{pmatrix} b_{11} & \ldots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \ldots & b_{nn} \end{pmatrix}.$$

Assume that ε is chosen such that

$$0 < \varepsilon < \frac{1}{|b_{i1}| + \cdots + |b_{in}|}$$

for every $i = 1, \ldots, n$. Note that here we clearly have

$$|b_{i1}| + \cdots + |b_{in}| \neq 0 \quad (i = 1, \ldots, n),$$

as the matrix B is invertible, too. Then the assertion

$$B \begin{pmatrix} u_1 - v_1 \\ \vdots \\ u_n - v_n \end{pmatrix} = \begin{pmatrix} x_1 - y_1 \\ \vdots \\ x_n - y_n \end{pmatrix}$$

implies that

$$|x_i - y_i| = |b_{i1}(u_1 - v_1) + \cdots + b_{in}(u_n - v_n)| \leq$$
$$\leq \varepsilon(|b_{i1}| + \cdots + |b_{in}|) < 1 \quad (i = 1, \ldots, n).$$

As $x_i, y_i \in \mathbb{Z}$ $(i = 1, \ldots, n)$, from this we get

$$|x_i - y_i| = 0 \quad (i = 1, \ldots, n).$$

This proves the statement. □

Remark 8.1.1 We mention without a proof that the above two properties already characterize the lattices in $\mathbb{R}^n$: a non-empty subset of $\mathbb{R}^n$ is a lattice if and only if it is an Abelian group (with respect to addition).

Remark 8.1.2 Observe that with the choice $a_1' = (1, 0)$, $a_2' = (1, 1)$ we obtain

$$\Lambda' = \{x_1 a_1' + x_2 a_2' : x_1, x_2 \in \mathbb{Z}\} = \mathbb{Z}^2,$$

as well. This shows that a lattice can be generated by several vector systems. So the question naturally arises: how can we characterize the vector systems generating the same lattice?

In what follows, we shall be concerned with the above question. For this, we shall need a new notion.

Definition 8.1.3 *Let A be a quadratic real matrix. If the entries of A are integers and* $\det(A) = \pm 1$*, then we say that A is a unimodular matrix.*

By the help of the above notion we can answer our previous question. In fact we could do this in full generality, however, for simplicity we restrict ourselves to the case of full lattices. (It will be sufficient also for our later purposes.)

Theorem 8.1.3 *Let $a_1, \dots, a_n$ and $b_1, \dots, b_n$ be two systems of linearly independent vectors in $\mathbb{R}^n$. Then the lattices generated by them coincide if and only if the basis transformation matrix between these two vector systems, as bases of $\mathbb{R}^n$, is unimodular.*

Proof. Denote by S the basis transformation matrix, that is, let S be that $n \times n$ real matrix, for which

$$S^T \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \tag{8.1}$$

holds. (Here S^T, as usual, stands for the transpose of S.) Further, let Λ_1 and Λ_2 be the lattices generated by the vectors $a_1, \dots, a_n$ and $b_1, \dots, b_n$, respectively.

Assume first that S is a unimodular matrix. Observe that then S^T is also unimodular. Thus by (8.1) we immediately obtain that $b_1, \dots, b_n \in \Lambda_1$, whence

$$\Lambda_2 \subseteq \Lambda_1$$

follows. On the other hand, as S (whence also S^T) is invertible, by (8.1) we get

$$\left(S^T\right)^{-1} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Since obviously $\left(S^T\right)^{-1}$ is also unimodular, hence we get $a_1, \dots, a_n \in \Lambda_2$, and then

$$\Lambda_1 \subseteq \Lambda_2.$$

From these we immediately obtain that

$$\Lambda_1 = \Lambda_2$$

holds.

Assume now that

$$\Lambda_1 = \Lambda_2,$$

and write S for the basis transformation matrix. Then, since $b_1, \ldots, b_n \in \Lambda_1$ (in view of the assertion (8.1)) we conclude that S^T has integer entries. Similarly, as $a_1, \ldots, a_n \in \Lambda_2$, we also obtain that $\left(S^T\right)^{-1}$ has integer entries, as well. Therefore the matrices S and S^{-1} also have integer entries. Hence in particular, $\det(S)$ and $\det(S^{-1})$ are integers. On the other hand, by the multiplicative property of determinants we get

$$\det(S) \cdot \det(S^{-1}) = 1.$$

However, this implies that

$$\det(S) = \det(S^{-1}) = \pm 1.$$

So S is a unimodular matrix, which proves the theorem. $\square$

Based upon the above theorem, we may introduce an important invariant of a lattice Λ.

Definition 8.1.4 *Let Λ be a full lattice in $\mathbb{R}^n$, and let $a_1, \ldots, a_n$ be a generating system of Λ. Then the quantity*

$$\det(\Lambda) = |\det(a_1, \ldots, a_n)|$$

is called the lattice determinant of Λ. Here $(a_1, \ldots, a_n)$ is an $n \times n$ matrix, with row vectors $a_1, \ldots, a_n$.

Remark 8.1.3 In view of the previous theorem, $\det(\Lambda)$ is an invariant of the lattice, indeed: its value does not depend on the choice of the generating system $a_1, \ldots, a_n$, only on the lattice itself.

Example 8.1.2 Let $a_1 = (1, 0)$, $a_2 = (0, 1)$ and $b_1 = (1, 0)$, $b_2 = (1, 1)$. As we have already seen, these pairs of vectors generate the same full lattice

$$\Lambda = \mathbb{Z}^2$$

in $\mathbb{R}^2$. The lattice determinant of Λ is given by

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = \det(\Lambda) = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix}.$$

We give a nice geometric interpretation of the lattice determinant. For this we shall need the following notion.

Definition 8.1.5 *Let Λ be a full lattice in $\mathbb{R}^n$, and let $a_1, \ldots, a_n$ be a generating system of Λ. Then the set*

$$P = \{\lambda_1 a_1 + \cdots + \lambda_n a_n : 0 \leq \lambda_i < 1 \ (i = 1, \ldots, n)\}$$

is called a fundamental parallelepiped (or when $n = 2$, a fundamental parallelogram) of Λ.

The fundamental parallelepipeds hold several important properties. In what follows, we shall discuss some of them. The first property is just the geometric interpretation of the lattice determinant.

Theorem 8.1.4 *Let P be a fundamental parallelepiped of a full lattice Λ in $\mathbb{R}^n$. Then*

$$\det(\Lambda) = V(P),$$

where $V(P)$ denotes the volume (i.e., the n-dimensional Lebesgue measure) of P.

Proof. Let $a_1, \ldots, a_n$ be the basis of Λ with which P was defined, and let

$$a_i = (a_{1i}, \ldots, a_{ni}) \quad (i = 1, \ldots, n).$$

Clearly, the volume of P is given by the following n-dimensional integral:

$$\mathrm{V}(P) = \int_P \ldots \int 1 \, \mathrm{dx}_1 \ldots \mathrm{dx}_n.$$

Let T be the n-dimensional unit cube, that is

$$T = \{(y_1, \ldots, y_n) \in \mathbb{R}^n : 0 \leq y_i < 1 \ (i = 1, \ldots, n)\}.$$

Consider the transformation

$$x = Ay \quad (x, y \in \mathbb{R}^n)$$

on $\mathbb{R}^n$, where

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix}.$$

The Jacobian determinant of the transformation is $\det(A)$. Observe that the image of the unit cube T under this transformation is just the fundamental parallelepiped P - this immediately follows from the definition of P. Thus by the rule of n-dimensional integral with substitution we obtain

$$\mathrm{V}(P) = \int_T \dots \int |\det(A)| \, \mathrm{dy}_1 \dots \mathrm{dy}_n =$$

$$= |\det(A)| \int_0^1 \dots \int_0^1 1 \, \mathrm{dy}_1 \dots \mathrm{dy}_n = |\det(A)| = \det(\Lambda),$$

which proves our theorem. □

The next theorem shows an important property of the fundamental parallelepipeds, namely that their shifts provide a tiling of $\mathbb{R}^n$.

Theorem 8.1.5 *Let P be a fundamental parallelepiped of a full lattice Λ in $\mathbb{R}^n$. Then the shifts of P with lattice vectors are disjoint, that is, for any $a, b \in \Lambda$ with $a \neq b$ we have*

$$(P + a) \cap (P + b) = \emptyset.$$

Further, the shifts of P with lattice vectors cover $\mathbb{R}^n$, that is, for any vector $x \in \mathbb{R}^n$ there exists (by the previous property, uniquely determined) $a \in \Lambda$, such that

$$x \in P + a$$

holds. Here we use the standard notation

$$P + a = \{p + a : p \in P\}.$$

Proof. To prove the first statement, assume that there is an $x \in \mathbb{R}^n$ such that

$$x \in (P + a) \cap (P + b)$$

holds with some $a, b \in \Lambda$. We show that then necessarily $a = b$, which will verify our claim. The above assertion yields that

$$p + a = q + b$$

with some $p, q \in P$. After rearrangement, this gives

$$a - b = p - q.$$

Let $a_1, \dots, a_n$ be the generating system of Λ which was used to define P. Observe that $a_1, \dots, a_n$ is a basis of $\mathbb{R}^n$. We can write the above identity in the form

$$(u_1 - v_1)a_1 + \cdots + (u_n - v_n)a_n = (x_1 - y_1)a_1 + \cdots + (x_n - y_n)a_n,$$

where

$$u_1, \dots, u_n \quad \text{and} \quad v_1, \dots, v_n$$

are the integer coefficients of a and b, respectively, with respect to the system $a_1, \dots, a_n$, while

$$x_1, \dots, x_n \quad \text{and} \quad y_1, \dots, y_n$$

are the coefficients from the interval $[0, 1)$ of p and q, respectively, in the same basis. Comparing the coefficients we immediately obtain that

$$u_1 = v_1, \dots, u_n = v_n \quad \text{and} \quad x_1 = y_1, \dots, x_n = y_n.$$

Hence

$$a = b \quad \text{and} \quad p = q,$$

and our claim follows.

To prove the second statement of the theorem let $x \in \mathbb{R}^n$ be arbitrary, and consider again the vectors $a_1, \dots, a_n$ defining P (which generate Λ and form a basis of $\mathbb{R}^n$ at the same time). Let

$$x = x_1 a_1 + \cdots + x_n a_n \quad (x_1, \dots, x_n \in \mathbb{R}),$$

and denote by y_i the integer part of x_i, and by λ_i the fractional part of x_i $(i = 1, \dots, n)$; that is,

$$0 \le \lambda_i < 1, \quad y_i \in \mathbb{Z}, \quad x_i = y_i + \lambda_i \quad (i = 1, \dots, n).$$

Thus using the notation

$$a = y_1 a_1 + \cdots + y_n a_n$$

and

$$p = \lambda_1 a_1 + \dots \lambda_n a_n,$$

we have

$$x = a + p,$$

that is

$$x \in P + a.$$

Hence the theorem is proved. $\square$

8.2 The theorem of Minkowski

In this section we discuss an important theorem, namely, the theorem of Minkowski. This result (as we shall see later) is an important tool in the investigation of many problems of different types.

To prove the theorem, we shall need another famous theorem due to Blichfeldt, so we start with this theorem.

Theorem 8.2.1 (Theorem of Blichfeldt) *Let Λ be a full lattice in $\mathbb{R}^n$ and T be a Lebesgue-measurable set in $\mathbb{R}^n$, with volume $V(T)$ for which*

$$V(T) > \det(\Lambda)$$

is valid. Then there exist $x_1, x_2 \in T$, $x_1 \neq x_2$ such that

$$x_1 - x_2 \in \Lambda$$

holds.

Proof. Let P be a fundamental parallelepiped of Λ, and let P_a be the shift of P by a ($a \in \Lambda$), that is,

$$P_a := P + a \quad (a \in \Lambda).$$

Consider the intersections of these sets with T, and shift these intersections to P, that is, set

$$T_a = (T \cap P_a) - a \quad (a \in \Lambda).$$

Observe that

$$T_a = \{x \in P : x + a \in T\} \quad (a \in \Lambda),$$

hence in particular,

$$T_a \subseteq P.$$

However, then using that the volume (the n-dimensional Lebesgue-measure) is shift-invariant, we obtain that

$$\mathrm{V}(P) = \det(\Lambda) < \mathrm{V}(T) = \sum_{a \in \Lambda} \mathrm{V}(T_a).$$

Thus necessarily there exist $a_1, a_2 \in \Lambda$ such that the intersection of T_{a_1} and T_{a_2} is non-empty, that is

$$x \in T_{a_1} \cap T_{a_2}$$

holds for some $x \in P$. Therefore by putting

$$x_1 = x + a_1$$

and

$$x_2 = x + a_2$$

we get that

$$x_1, x_2 \in T$$

and

$$x_1 - x_2 = a_1 - a_2 \in \Lambda.$$

This proves our claim. $\square$

Our next theorem is the important Minkowski theorem. For its formulation we need to introduce (or rather recall) two notions.

Definition 8.2.1 *Let S be a subset of $\mathbb{R}^n$. If for any $x \in S$ we have $-x \in S$, then S is called centralsymmetric.*

Definition 8.2.2 *Let S be a subset of $\mathbb{R}^n$. If $x, y \in S$ and for any $\lambda \in [0,1]$ we also have $\lambda x + (1-\lambda)y \in S$, then we say that S is convex.*

Now we are in a position that we can formulate Minkowski's famous theorem concerning lattice points of convex bodies.

Theorem 8.2.2 (Minkowski theorem) *Let Λ be a full lattice in $\mathbb{R}^n$. Further, let S be a convex, centralsymmetric set in $\mathbb{R}^n$, holding one of the next two properties:*

i) $V(S) > 2^n \det(\Lambda)$,

ii) $V(S) = 2^n \det(\Lambda)$ *and S is compact.*

Then S contains a lattice point of Λ different from the origin.

Proof. First we prove the conclusion under the assumption i). Let

$$T = \frac{1}{2}S = \left\{\frac{1}{2}s : s \in S\right\}.$$

Then by a well-known property of the n-dimensional volume (Lebesgue-measure) we have

$$\mathrm{V}(T) = \frac{1}{2^n}\mathrm{V}(S) > \det(\Lambda).$$

Thus by the theorem of Blichfeldt there exist $t_1, t_2 \in T$, $t_1 \neq t_2$ for which

$$t_1 - t_2 \in \Lambda.$$

Hence by the definition of the set T there exist points s_1, s_2 in S for which

$$\frac{1}{2}s_1 - \frac{1}{2}s_2 \in \Lambda$$

holds. As $s_1 \neq s_2$, the above lattice point of Λ is different from the origin. On the other hand, by the centralsymmetry and convexity of S we obtain

$$\frac{1}{2}s_1 - \frac{1}{2}s_2 = \frac{1}{2}s_1 + \frac{1}{2}(-s_2) \in S,$$

as well. Hence S contains a lattice point of Λ different from the origin, indeed. This proves the theorem in this case.

Now we prove that the conclusion also holds assuming condition ii). For this, introduce the following notation:

$$S_n := \left(1 + \frac{1}{n}\right) S = \left\{\left(1 + \frac{1}{n}\right) s : s \in S\right\} \quad (n \in \mathbb{N}).$$

Observe that for every $n \in \mathbb{N}$ the set S_n is a convex, centralsymmetric set, with volume $\mathrm{V}(S_n) > 2^n \det(\Lambda)$. Thus condition i) is valid for the sets S_n $(n \in \mathbb{N})$, so there exist lattice points $x_1, x_2, \ldots$ in Λ different from the origin, with

$$x_n \in S_n \quad (n \in \mathbb{N}).$$

As by the definition of the sets S_n $(n \in \mathbb{N})$

$$S_1 \supset S_2 \supset S_3 \supset \ldots$$

holds, we get that

$$x_n \in S_1 \quad (n \in \mathbb{N})$$

is also valid. Since by our assumption S is compact, thus obviously S_1 is compact, as well. However, then if the set

$$X := \{x_n : n \in \mathbb{N}\}$$

would be infinite, then (since it is a subset of the bounded set S_1) by the Bolzano-Weierstrass theorem it would have an accumulation point. However, this by $X \subseteq \Lambda$ is not possible, as Λ is discrete. Thus X is necessarily a finite set. But then for some index i_0 we have

$$x_i = x_{i_0}$$

for all $i \in I$, where I is an infinite subset of $\mathbb{N}$. Then

$$x_{i_0} \in S_i \quad (i \in I).$$

On the other hand, by the compactness of S we simply obtain that

$$\bigcap_{i \in I} S_i = S.$$

Indeed,

$$\bigcap_{i \in I} S_i \supseteq S$$

trivially holds. Hence if

$$\bigcap_{i \in I} S_i \neq S$$

would be valid, then there would be an

$$x \in \bigcap_{i \in I} S_i$$

such that

$$x \notin S.$$

However, then using that S is closed, the distance of x and S is positive, that is

$$\inf_{y \in S} |x - y| > 0.$$

However, this by the definition of the sets S_n $(n \in \mathbb{N})$ is clearly impossible.

From the above assertions we obtain that $x_{i_0} \in S$, which proves the theorem. □

Remark 8.2.1 One can easily check that all assumptions of the Minkowski theorem are necessary, omitting any of them the statement is not valid any more. The analysis of this question is left to the Reader, among the Exercises.

Finally, as an application of the Minkowski theorem we show that certain systems of linear inequalities always have non-trivial solutions. The next theorem is often called as Minkowski's linear form theorem.

Theorem 8.2.3 *Let α_{ij} $(i,j=1,\dots,n)$ be real numbers,*

$$A=\begin{pmatrix}\alpha_{11} & \dots & \alpha_{1n}\\ \vdots & \ddots & \vdots\\ \alpha_{n1} & \dots & \alpha_{nn}\end{pmatrix},$$

and assume that $\det(A)\neq 0$. *Let $c_1,\dots,c_n$ be positive real numbers with*

$$c_1\dots c_n\geq|\det(A)|.$$

Then the system of inequalities

$$|L_1(x)|\leq c_1,\ |L_2(x)|\leq c_2,\ \dots,\ |L_n(x)|\leq c_n, \tag{8.2}$$

where

$$L_i(x)=\alpha_{i1}x_1+\dots+\alpha_{in}x_n$$

and $x=(x_1,\dots,x_n)$ is an unknown tuple in $\mathbb{R}^n$, always has a non-trivial (not identically zero) solution $(x_1,\dots,x_n)\in\mathbb{Z}^n$.

Proof. Let $\Lambda=\mathbb{Z}^n$. Then clearly, Λ is a full lattice in $\mathbb{R}^n$, with lattice determinant $\det(\Lambda)=1$. Let S denote the solution set of the inequality system (8.2). Obviously, S is a compact, convex, centralsymmetric set, with volume

$$\mathrm{V}(S)=\int_S\dots\int 1\ \mathrm{dx}_1\dots\mathrm{dx}_n.$$

(The integral above is of course n-dimensional.) Thus with the substitution

$$\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=A\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix},$$

by the rule of n-dimensional integral with substitution (using that the Jacobian is $\det(A^{-1})$), we obtain that

$$\mathrm{V}(S)=\int_{|y_1|\leq c_1}\dots\int_{|y_n|\leq c_n}|\det(A^{-1})|\ \mathrm{dy}_1\dots\mathrm{dy}_n=$$
$$=|\det(A^{-1})|\cdot 2^n\cdot c_1\cdots c_n\leq 2^n=2^n\det(\Lambda).$$

Hence our statement follows from case ii) of the Minkowski theorem. □

Remark 8.2.2 We mention that if in place of the inequality system in the theorem, we use the more restrictive system

$$|L_1(x)| \leq c_1,\ |L_2(x)| < c_2,\ \ldots,\ |L_n(x)| < c_n,$$

the conclusion of the statement remains valid. However, here the inequality $|L_1(x)| \leq c_1$ cannot be replaced by the strict inequality $|L_1(x)| < c_1$.

8.3 Exercises

Exercise 8.3.1 Let Λ be a full lattice in $\mathbb{R}^2$.

a) Prove that there are infinitely many lines which do not contain any lattice point from Λ.

b) Prove that there are infinitely many lines which contain precisely one lattice point from Λ.

c) Prove that if a line e contains two lattice points from Λ, then e contains infinitely many lattice points from Λ.

Exercise 8.3.2 Let Λ be a full lattice in $\mathbb{R}^2$. A line is called lattice line, if it contains at least two lattice points from Λ. (By the previous exercise we know that then this line contains infinitely many lattice points from Λ.) Prove the following statements!

a) The distances of any two neighboring lattice points of a lattice line are equal.

b) A line parallel to a lattice line which contains a lattice point is also a lattice line.

c) There exists a positive constant c depending only on Λ, such that the distance of any two parallel lattice lines is greater than c.

Exercise 8.3.3 Consider the following lattices Λ_1 and Λ_2 in $\mathbb{R}^2$, generated by the vectors a_1, a_2 and b_1, b_2, respectively:

a) $a_1 = (1, 0)$, $a_2 = (0, 1)$ and $b_1 = (1, 2)$, $b_2 = (3, 5)$,

b) $a_1 = (3, 4)$, $a_2 = (-1, 7)$ and $b_1 = (-2, 3)$, $b_2 = (-3, 2)$,

c) $a_1 = (1, 0)$, $a_2 = (0, 1)$ and $b_1 = (1, 2)$, $b_2 = (0, -1)$,

d) $a_1 = (1, 1)$, $a_2 = (-1, 1)$ and $b_1 = (1, -2)$, $b_2 = (2, 1)$,

e) $a_1 = (7, 6)$, $a_2 = (1, -1)$ and $b_1 = (-7, 6)$, $b_2 = (-1, -1)$,

f) $a_1 = (2, 3)$, $a_2 = (-1, 4)$ and $b_1 = (11, 1)$, $b_2 = (0, 1)$.

Decide that in which cases the lattices Λ_1 and Λ_2 are equal!

Exercise 8.3.4 Consider the following lattices Λ_1 and Λ_2 in $\mathbb{R}^3$, generated by the vectors a_1, a_2, a_3 and b_1, b_2, b_3, respectively:

a) $a_1 = (1, 0, 0)$, $a_2 = (0, 1, 0)$, $a_3 = (0, 0, 1)$ and
$b_1 = (1, 2, 0)$, $b_2 = (3, 5, 0)$ $b_3 = (1, 1, -1)$,

b) $a_1 = (-1, -1, 1)$, $a_2 = (-1, 3, 2)$, $a_3 = (-2, 2, 1)$ and
$b_1 = (-2, 3, 1)$, $b_2 = (-3, 7, 0)$ $b_3 = (2, 1, 0)$,

c) $a_1 = (1, 0, 0)$, $a_2 = (0, 1, 0)$, $a_3 = (0, 0, 1)$ and
$b_1 = (1, 2, 3)$, $b_2 = (0, -1, 7)$, $b_3 = (0, 0, -1)$,

d) $a_1 = (1, 1, 2)$, $a_2 = (-1, 1, 0)$, $a_3 = (2, 3, 1)$ and
$b_1 = (1, -2, 4)$, $b_2 = (2, 1, 5)$, $b_3 = (3, 0, -2)$,

e) $a_1 = (1, 2, 3)$, $a_2 = (-1, 1, 0)$, $a_3 = (0, 0, 1)$ and
$b_1 = (1, 3, 4)$, $b_2 = (1, 0, -7)$, $b_3 = (0, 0, -1)$,

f) $a_1 = (-1, 2, 3)$, $a_2 = (1, 1, 0)$, $a_3 = (0, 0, 1)$ and
$b_1 = (-1, 3, 4)$, $b_2 = (-1, 0, -7)$, $b_3 = (0, 0, -1)$.

Decide that in which cases the lattices Λ_1 and Λ_2 are equal!

Exercise 8.3.5 Calculate the lattice determinants of the lattices given in the previous two exercises!

Exercise 8.3.6 Let Λ be the lattice generated by the vectors $a = (2, 3)$, $b = (-1, 2)$ in $\mathbb{R}^2$, and let P be the fundamental parallelogram defined by a, b.

a) Give the area of P.

b) Let $x = (1, 0)$ and $y = (11, 13)$. Decide whether these vectors are contained in the lattice Λ or not.

c) Let $u = (1, 1)$. Give all vectors $v \in \mathbb{R}^2$, for which $u - v$ is in the lattice Λ.

d) Let $T = [0, 2] \times [0, 3]$. Prove that there exist vectors $u, v \in T$ for which we have $u - v \in \Lambda$.

Exercise 8.3.7 Let Λ be the lattice generated by the vectors $a = (1, 1, 1)$, $b = (0, 2, 1)$, $c = (0, 0, 3)$ in $\mathbb{R}^3$, and let P be the fundamental parallelepiped of the lattice generated by a, b, c.

a) Give the volume of P.

b) Let $x = (1, 0, -1)$ and $y = (1, 3, -1)$. Decide whether these vectors are contained in the lattice Λ or not.

c) Let $u = (1, 1, 1)$. Give all vectors $v \in \mathbb{R}^3$, for which $u - v$ is in the lattice Λ.

Exercise 8.3.8 Characterize those vectors $a \in \mathbb{Z}^2$, for which one can find a vector $b \in \mathbb{Z}^2$ such that a, b generates the lattice $\Lambda = \mathbb{Z}^2$ in $\mathbb{R}^2$.

Exercise 8.3.9 As a generalization of the previous problem, characterize those vectors $a \in \mathbb{Z}^n$, for which one can find vectors $b_2, \ldots, b_n \in \mathbb{Z}^n$ such that $a, b_2, \ldots, b_n$ generates the lattice $\Lambda = \mathbb{Z}^n$ in $\mathbb{R}^n$.

Exercise 8.3.10 Prove that the assumptions in the Minkowski theorem are all necessary. For this, take $\Lambda = \mathbb{Z}^2$ and give an example for a set $S \subseteq \mathbb{R}^2$ for which one of the following holds:

a) S is convex and $\mathrm{V}(S) > 4$,

b) S is centralsymmetric and $\mathrm{V}(S) > 4$,

c) S is convex, centralsymmetric and $\mathrm{V}(S) = 4$,

however, S does not contain any lattice point of Λ different from the origin!

Exercise 8.3.11 Prove that it is possible visit all squares of an infinite chessboard with a knight, that is, from any square we can get to any other square by valid moves!

Chapter 9

Waring's problem

In this chapter we discuss representations of positive integers as sums of k-th powers.

9.1 Basic notions

The basic question of the topic is due to Waring: is it true, that for any exponent $k \geq 2$ one can find a positive integer $g(k)$, such that every positive integer can be represented as the sum of $g(k)$ k-th powers?

To be precise, we formulate the following definition.

Definition 9.1.1 *Let $k \geq 2$, and let $G(k)$ be the smallest positive integer, for which every positive integer can be represented as the sum of $G(k)$ k-th power. Here by a k-th power we mean the k-th power of a non-negative integer.*

Remark 9.1.1 Though our purpose is the study of the representations of natural numbers (positive integers), it is still worth to consider the number $0 = 0^k$ to be a k-th power, as well. Indeed, otherwise we would have problems with trying to represent the positive integers as the sum of k-th powers with the same number of terms. But in this way we get a problem which can be nicely formulated and easily understood.

It is also important to note that at this point we do not yet know that $g(k)$, hence $G(k)$ exists for $k \geq 2$. In what follows, we shall see that in fact $g(k)$, and hence also $G(k)$ exists.

9.2 Representation of positive integers as sums of squares

In this section we study Waring's problem for $k = 2$. We make a detailed analysis of representations of positive integers as sums of squares: we study separately the cases of representability with sums one, two, three and four squares. Clearly, the numbers being representable as a 'sum of one square' are just the squares. Thus the first interesting case is to describe those positive integers which can be represented as a sum of two squares. To answer this question, we shall need the following result due to Euler.

Lemma 9.2.1 *Every prime of the form $4k + 1$ can be represented as a sum of two squares.*

Proof. Let p be any prime of the form $4k + 1$. We know that then

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = 1$$

holds, that is, there exists an integer a for which

$$a^2 \equiv -1 \pmod{p}.$$

Take such an integer a, and let $b_1 = (p, 0)$, $b_2 = (a, 1)$. Clearly, b_1, b_2 are linearly independent vectors in $\mathbb{R}^2$. Consider the full lattice Λ in $\mathbb{R}^2$ generated by them. We get that the lattice determinant of Λ is given by

$$\det(\Lambda) = \begin{vmatrix} p & 0 \\ a & 1 \end{vmatrix} = p.$$

Also observe that by $b_1, b_2 \in \mathbb{Z}^2$ we have $\Lambda \subseteq \mathbb{Z}^2$.

Take the set

$$S = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq \frac{4p}{\pi}\right\}$$

in $\mathbb{R}^2$. In fact, S is the closed disc of radius $\sqrt{4p/\pi}$ centered in the origin. That is, S is a compact, convex, centralsymmetric set, of area

$$V(S) = 4p.$$

Thus by part ii) of the Minkowski theorem, we obtain that S contains a lattice point (u, v) in Λ, different from the origin. In particular, here we have $u, v \in \mathbb{Z}$. Further, $(u, v) \in \Lambda$ implies that

$$(u, v) = x_1(p, 0) + x_2(a, 1)$$

holds with some integers x_1, x_2. From this we get that

$$u^2 + v^2 = x_1^2 p^2 + 2pax_1x_2 + x_2^2(a^2 + 1).$$

Hence by the choice of a we infer that

$$p \mid u^2 + v^2.$$

On the other hand, $(u, v) \in S$, $(u, v) \neq (0, 0)$ yields

$$0 < u^2 + v^2 \leq \frac{4p}{\pi} < 2p.$$

Thus necessarily

$$u^2 + v^2 = p,$$

which proves our claim. □

Remark 9.2.1 As a trivial, but important observation note that

$$2 = 1^2 + 1^2,$$

that is, the prime 2 can be represented as the sum of two squares.

The next theorem of Euler gives a precise description of positive integers being representable as the sum of two squares. In view of the trivial representation

$$1 = 1^2 + 0^2,$$

it is sufficient to study the representability of integers greater than one.

Theorem 9.2.1 (Euler's theorem) *Let $n \in \mathbb{N}$, $n > 1$. Then n can be represented as the sum of two squares, if and only if in the prime factorization of n the exponent of each prime of the form $4k - 1$ is even.*

Proof. First we show that the set of numbers being representable as sums of two squares is closed with respect to multiplication. (That is, this set is a commutative (multiplicative) semigroup with unit element.) For this, let m and n be two such numbers, with representations

$$m = x_1^2 + x_2^2$$

and

$$n = y_1^2 + y_2^2.$$

Then

$$mn = (x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 + x_2y_2)^2 + (x_1y_2 - x_2y_1)^2$$

holds, which proves our claim.

We already now that 2, as well as all primes of the form $4k + 1$ can be represented as sums of two squares. The trivial representation

$$q^2 = q^2 + 0^2$$

shows that the same is true for the squares of primes q of the form $4k - 1$, as well. From this by our previous observation it follows that the numbers of the shape given in the statement can be represented as sums of two squares, indeed. Thus the condition in the theorem is sufficient.

To prove the necessity of the condition, assume to the contrary that a positive integer n can be represented as

$$n = x^2 + y^2$$

with some non-negative integers x, y, however, there is a prime q of the form $4k - 1$ occurring in the prime factorization of n in an odd exponent. Clearly, we have $xy \neq 0$. Write α, β, γ for the exponents of q in the prime factorizations of x, y, n, respectively. These exponents are non-negative integers, and by our assumption, γ is odd. Then by the notation

$$x = q^\alpha x_0, \quad y = q^\beta y_0, \quad n = q^\gamma n_0$$

where x_0, y_0, n_0 are positive integers such that

$$q \nmid x_0 y_0 n_0,$$

we obtain

$$q^{2\alpha}x_0^2 + q^{2\beta}y_0^2 = q^{\gamma}n_0. \tag{9.1}$$

Observe that after cancellation with the highest possible power of q, that is, with

$$q^{\min(2\alpha,2\beta,\gamma)},$$

there can be at most one term which is divisible by q. Indeed, three terms clearly cannot remain divisible by q, and if there remains two terms divisible by q, then necessarily the third term also would be divisible by q. However, as γ is odd, this is possible only if $\alpha = \beta$, and

$$\alpha = \beta = \min(2\alpha, 2\beta, \gamma) < \gamma.$$

Thus from equation (9.1) we obtain

$$x_0^2 + y_0^2 = q^{\gamma-2\alpha}n_0.$$

In particular, $\gamma > 2\alpha$ implies

$$x_0^2 + y_0^2 \equiv 0 \pmod{q}. \tag{9.2}$$

As $q \nmid x_0$, the integer x_0 is invertible modulo q. That is, there exists an integer z_0, such that

$$x_0 z_0 \equiv 1 \pmod{q}$$

holds. Thus multiplying the congruence (9.2) with z_0^2, after rearrangement we obtain

$$(y_0 z_0)^2 \equiv -1 \pmod{q}.$$

Therefore -1 is a quadratic residue modulo q, that is,

$$\left(\frac{-1}{q}\right) = 1.$$

However, this contradicts the fact that q is a prime of the form $4k-1$, and hence

$$\left(\frac{-1}{q}\right) = (-1)^{\frac{q-1}{2}} = -1.$$

This contradiction proves the theorem. □

We conclude the section with giving two theorems, without proofs. The first one is an important result of Gauss, which describes integers representable as sums of three squares.

Theorem 9.2.2 (Gauss theorem) *A positive integer n is* not *representable as the sum of three squares, if and only if n is of the shape*

$$n = 4^a(8b + 7)$$

with some non-negative integers a, b.

The last theorem of the section, which is due to Lagrange, shows that every positive integer can be represented as a sum of four squares.

Theorem 9.2.3 (Lagrange theorem) *Every positive integer can be represented as a sum of four squares.*

Remark 9.2.2 In view of the above theorems, we see that $g(2)$, whence also $G(2)$ exists, and $G(2) = 4$.

9.3 Representation of positive integers as sums of higher powers

In this section we give a theorem of Hilbert, which provides an answer to Waring's question.

Theorem 9.3.1 (Hilbert theorem) *For every $k \geq 2$ $g(k)$, and hence also $G(k)$ exists.*

Remark 9.3.1 The value of $G(k)$ is known for certain small values of k:

$$G(2) = 4, \quad G(3) = 9, \quad G(4) = 19.$$

In the theory discussed in the previous section we know that $G(2) = 4$ indeed holds. By representing 23 as sums of cubes and 79 as sums of fourth powers, one can easily check that $G(3) \geq 9$ and $G(4) \geq 19$, respectively.

9.4 Exercises

Exercise 9.4.1 Determine which of the following numbers can be represented as sums of two squares:

- 90
- 120
- 121
- 450
- 1111
- 7000000
- 123456789012

Exercise 9.4.2 Prove that the equation

$$x^2 + y^2 = 3z^2$$

has no solutions in positive integers x, y, z.

Exercise 9.4.3 Let $a, b \in \mathbb{N}$ such that both

$$x^2 + y^2 = az^2$$

and

$$x^2 + y^2 = bz^2$$

are solvable in positive integers x, y, z. Prove that then the equation

$$x^2 + y^2 = abz^2$$

is also solvable in positive integers x, y, z.

Exercise 9.4.4 Prove that for any $k \in \mathbb{N}$ one can find an $n \in \mathbb{N}$ such that none of the numbers

$$n + 1, \ldots, n + k$$

can be represented as a sum of two squares!

Exercise 9.4.5 Prove that for any $k \in \mathbb{N}$ one can find an $n \in \mathbb{N}$, such that n can be represented as a sum of two squares, however, none of the numbers

$$n-k, \ldots, n-1, n+1, \ldots, n+k$$

can be represented as a sum of two squares!

Exercise 9.4.6 Determine which of the following numbers can be represented as sums of three squares:

- 77
- 799
- 120
- 350
- 1111
- 80000
- 12340005

Exercise 9.4.7 Prove that 130 cannot be written as a sum of three squares!

Exercise 9.4.8 Give an example for $a, b \in \mathbb{N}$ which can be represented as sums of three squares, however ab cannot be written as a sum of three squares!

Exercise 9.4.9 Let $k \in \mathbb{N}$ be arbitrary. Prove that if n can be written as a sum of three squares, then so is n^k.

Exercise 9.4.10 Prove that for any $n > 1$ one of the numbers $n-1$ and $n+1$ can be written as a sum of three squares!

Exercise 9.4.11 Prove that the equation

$$x^2 + y^2 + z^2 + x + y + z = 1$$

is not solvable in rational numbers x, y, z.

Exercise 9.4.12 Prove that there are infinitely many positive integers which can be represented as a sum of three squares, but cannot be represented as a sum of less than three squares!

Exercise 9.4.13 Prove that there are infinitely many positive integers, which can be represented as a sum of four squares only such that none of the squares is zero!

Chapter 10

Elements of algebraic number theory

In this chapter we outline the basics of algebraic number theory.

10.1 Algebraic and transcendental numbers

This section introduces the most important notions and assertions of algebraic number theory.

Definition 10.1.1 *The number $\alpha \in \mathbb{C}$ is algebraic, if there exists a not identically zero polynomial $p(x) \in \mathbb{Q}[x]$, such that α is a root of that. If such a polynomial does not exist, then α is transcendental. If $p(x)$ is a monic polynomial of minimal degree, then it is called the defining monic polynomial of α. If the defining monic polynomial of α is of degree n, then we say that α is an algebraic number of degree n. Notation:* $\deg(\alpha) = n$. *The roots of the defining monic polynomial of α are called the algebraic conjugates of α.*

Remark 10.1.1 At this point we mention two things.

1) A complex number α is an algebraic number of degree one if and only if α is rational. Indeed, if α is algebraic of degree one, then its defining monic polynomial is of the form

$$p(x) = x - r,$$

where $r \in \mathbb{Q}$. Thus $\alpha = r$ implies $\alpha \in \mathbb{Q}$. On the other hand, if $\alpha \in \mathbb{Q}$, then of course α is a root of the polynomial $x - \alpha \in \mathbb{Q}[x]$, hence it is an algebraic number of degree one.

2) For every $n \geq 1$ there exists an algebraic number of degree n. Namely, $\alpha = \sqrt[n]{2}$ is algebraic of degree n. This follows from the fact that α is a root of the polynomial

$$p(x) = x^n - 2,$$

so α is algebraic and $\deg(\alpha) \leq n$. At the same time, the degree of α cannot be smaller than n, since then the defining monic polynomial of α would divide $p(x)$ - but in view of the Schönemann-Eisenstein theorem $p(x)$ is irreducible in $\mathbb{Q}[x]$. Hence necessarily $\deg(\alpha) = n$. Our observation concerning irreducibility proves to be very important: our next theorem shows that this property characterizes the defining monic polynomials.

Theorem 10.1.1 *Let α be an algebraic number. Then the defining monic polynomial of α is irreducible over $\mathbb{Q}$. On the other hand, if α is a root of a monic, irreducible polynomial in $\mathbb{Q}[x]$, then it is the defining monic polynomial of α.*

Proof. Let $p(x) \in \mathbb{Q}[x]$ be the defining monic polynomial of α. We show that then $p(x)$ is irreducible over $\mathbb{Q}$. Assume to the contrary that it is not the case, and we can write

$$p(x) = f(x)g(x)$$

with some polynomials $f(x), g(x) \in \mathbb{Q}[x]$ of positive degrees. Then since

$$0 = p(\alpha) = f(\alpha)g(\alpha),$$

we have that

$$f(\alpha) = 0 \quad \text{or} \quad g(\alpha) = 0$$

holds. However, this contradicts the minimality of the degree of $p(x)$, which proves our statement.

Assume now that α is a root of an irreducible monic polynomial $f(x) \in \mathbb{Q}[x]$. Let $p(x)$ be the defining monic polynomial of α. Let

$$g(x) = \gcd(f(x), p(x))$$

in $\mathbb{Q}[x]$. We may assume that $g(x)$ is a monic polynomial. Since α is a root of both f and p, hence it is also a root of g, so $\deg(g(x)) \geq 1$. However, then by the irreducibility of f (which we assumed) and by the irreducibility of p (which follows from the first part of the theorem) we obtain

$$f(x) = g(x) = p(x),$$

which was to be proved. $\square$

Definition 10.1.2 *The algebraic number α is called an algebraic integer, if its defining monic polynomial belongs to $\mathbb{Z}[x]$.*

Remark 10.1.2 At this point we mention three things.

1) In some cases it might be disturbing that we use the expression 'integer' both for the algebraic integers, and for the usual integers in $\mathbb{Z}$. Thus to avoid misunderstandings, the elements of $\mathbb{Z}$ are often called rational integers.

2) The algebraic integers of degree one are precisely the rational integers. It can be easily checked in view of point 1) of our previous remark.

3) The algebraic number $\alpha = \sqrt[n]{2}$ is in fact an algebraic integer of degree n - this can also be readily checked.

In the following we shall see that it is not necessary to know the defining monic polynomial of an algebraic number α to be able to decide whether α is an algebraic integer or not: this property already follows from the fact that α is a root of a monic polynomial with integral coefficients. However, to prove this assertion, we shall need the following famous lemma, due to Gauss.

Lemma 10.1.1 (Gauss lemma) *Let $f, g \in \mathbb{Z}[x]$,*

$$f(x) = a_n x^n + \cdots + a_1 x + a_0,$$

$$g(x) = b_m x^m + \cdots + b_1 x + b_0,$$

where $a_n b_m \neq 0$. Assume that f and g are primitive, that is,

$$\gcd(a_n, \ldots, a_1, a_0) = \gcd(b_m, \ldots, b_1, b_0) = 1.$$

Then $f \cdot g$ is also primitive.

Proof. Let

$$h(x) = f(x)g(x),$$

and assume to the contrary that $h(x)$ is not primitive. Then there exists a prime p which divides all coefficients of h. Let i and j be maximal such that

$$p \mid a_n, a_{n-1}, \ldots, a_{i+1} \quad \text{but} \quad p \nmid a_i,$$

and

$$p \mid b_m, b_{m-1}, \ldots, b_{j+1} \quad \text{but} \quad p \nmid b_j.$$

Since f and g are primitive, such indices i and j exist. Then by our assumptions p divides the coefficient of x^{i+j} in $h(x)$, that is

$$p \mid \cdots + a_{i+1}b_{j-1} + a_ib_j + a_{i-1}b_{j+1} + \ldots.$$

However, then $p \mid a_ib_j$. This is a contradiction, and the lemma is proved. □

Theorem 10.1.2 *Assume that an algebraic number α is a root of a monic polynomial with integer coefficients. Then α is an algebraic integer.*

Proof. Assume that α is a root of a monic polynomial $f(x) \in \mathbb{Z}[x]$, and let $p(x) \in \mathbb{Q}[x]$ be the defining monic polynomial of α. Then, as α is a root of both f and p, we get that

$$\gcd(f(x), p(x))$$

in $\mathbb{Q}[x]$ is of degree at least one. However, then by the irreducibility of $p(x)$ this common divisor is $p(x)$ itself, whence

$$p(x) \mid f(x) \quad \text{over } \mathbb{Q}.$$

Thus we can write

$$f(x) = p(x)q(x)$$

with some polynomial $q(x) \in \mathbb{Q}[x]$. Let p_0 and q_0 be the minimal positive integers for which

$$p_0p(x), q_0q(x) \in \mathbb{Z}[x]$$

hold. Then the above polynomials are clearly primitive. But then by

$$p_0q_0f(x) = (p_0p(x))(q_0q(x))$$

using the Gauss lemma, we obtain that $p_0q_0f(x)$ is also primitive. However, this is only possible if

$$p_0 = q_0 = 1,$$

that is, $p(x) \in \mathbb{Z}[x]$, which proves the theorem. $\square$

The last theorem of the section is of great importance: it shows that the sets of algebraic numbers and algebraic integers form rich structures.

Theorem 10.1.3 *With respect to the usual addition and multiplication in $\mathbb{C}$, the set of algebraic numbers form a field, while the set of algebraic integers form a ring.*

Proof. Obviously, it is sufficient to check that our sets are closed under addition and multiplication, and (since 0 and 1 are algebraic numbers, further, algebraic integers) the appropriate inverses also belong to our sets. Indeed, the necessary properties of the operations clearly hold, since we do not leave the field of complex numbers.

Being closed with respect to the operations is only checked in case of the set of algebraic numbers and addition: all the other cases can be checked similarly. So let α and β be algebraic numbers. We show that then $\alpha + \beta$ is algebraic, as well. For this let $p(x)$ and $q(x)$ be the defining monic polynomial of α and β, respectively. Then we can write

$$p(x) = (x - \alpha_1) \dots (x - \alpha_n)$$

and

$$q(x) = (x - \beta_1) \dots (x - \beta_m),$$

where $n = \deg(\alpha)$, $m = \deg(\beta)$, and $\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_m$ are the algebraic conjugates of α and β, respectively. We may clearly assume that $\alpha = \alpha_1$ and $\beta = \beta_1$. Note that $\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_m$ are pairwise distinct, since $p(x)$ and $q(x)$ are irreducible over $\mathbb{Q}$. Let

$$h(x) = \prod_{i=1}^{n} \prod_{j=1}^{m} (x - \alpha_i - \beta_j).$$

Obviously $\alpha + \beta$ is a root of $h(x)$, so we only need to show that $h(x) \in \mathbb{Q}[x]$. For this observe that

$$h(x) = q(x - \alpha_1) \dots q(x - \alpha_n)$$

holds, that is, $h(x)$ is a symmetric polynomial of $\alpha_1, \ldots, \alpha_n$ over $\mathbb{Q}[x]$. Thus by the fundamental theorem of symmetric polynomials, $h(x)$ can be represented as an elementary symmetric polynomial of $\alpha_1, \ldots, \alpha_n$, with coefficients from $\mathbb{Q}[x]$. In other words, we have

$$h(x) = F(\sigma_1, \ldots, \sigma_n),$$

where F is a polynomial in n variables with coefficients from $\mathbb{Q}[x]$, while σ_i $(i = 1, \ldots, n)$ is the i-th symmetric polynomial of $\alpha_1, \ldots, \alpha_n$. Hence $\sigma_1, \ldots, \sigma_n$ are just the coefficients of $p(x)$ (up to signs), so they are rational numbers. This shows that $h(x) \in \mathbb{Q}[x]$, which proves our statement. □

10.2 Algebraic number fields

In this section we study the finite degree extensions of $\mathbb{Q}$. First we recall the notion of field extension.

Definition 10.2.1 *Let K and L be fields, $K \subseteq L$. Then we say that L is an extension of K. By the degree of the extension we mean the dimension of L, as a vector space over K.*

Now we are ready to introduce the notion of algebraic number fields.

Definition 10.2.2 *The finite degree extensions of $\mathbb{Q}$ are called algebraic number fields.*

Remark 10.2.1 It is well-known that any finite degree extension L of a field K can be obtained as a simple extension of K, that is, in the form $L = K(\alpha)$, where α is an algebraic element over K, and $K(\alpha)$ is the smallest field containing both K and α. In our case it means that the algebraic number fields can be obtained as $\mathbb{Q}(\alpha)$, where $\alpha \in \mathbb{C}$ is an algebraic number.

In the following we study algebraic number fields. Our first theorem is simple and natural.

Theorem 10.2.1 *All elements of an algebraic number field of degree n are algebraic numbers of degree at most n.*

Proof. Let K be an algebraic number field of degree n, and let $\alpha \in K$. Then the numbers

$$\alpha^n, \ldots, \alpha, 1$$

(being $n+1$ elements of an n-dimensional vector space) are necessarily linearly dependent over $\mathbb{Q}$. However, then there exist rational numbers $a_n, \ldots, a_1, a_0$, not all zero, such that

$$a_n\alpha^n + \cdots + a_1\alpha + a_0 = 0.$$

Therefore α is a root of the polynomial

$$a_n x^n + \cdots + a_1 x + a_0$$

of degree at most n from $\mathbb{Q}[x]$. Hence α is algebraic indeed, and of course $\deg(\alpha) \leq n$. $\square$

Our next theorem gives a precise description of the representations of the elements of algebraic number fields.

Theorem 10.2.2 *Let $K = \mathbb{Q}(\alpha)$ be an algebraic number field, where α is an algebraic number of degree n. Then*

$$K = \{a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 : a_{n-1}, \ldots, a_1, a_0 \in \mathbb{Q}\},$$

in particular, K is of degree n.

Proof. Let

$$T = \{a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 : a_{n-1}, \ldots, a_1, a_0 \in \mathbb{Q}\}.$$

$K = \mathbb{Q}(\alpha)$ implies that $T \subseteq K$. As $\alpha \in T$ and $\mathbb{Q} \subseteq T$, thus it is sufficient to show that $(T, +, \cdot)$ is a field. Since the identities for the operations trivially hold, further, $0, 1 \in T$, thus it is sufficient to check that T is closed with respect to addition, multiplication, and taking additive and multiplicative inverse. We check these properties separately.

The fact that T is closed with respect to addition, trivially follows from the definition of T.

Let now $t_1, t_2 \in T$,

$$t_1 = a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 \quad (a_{n-1}, \ldots, a_1, a_0 \in \mathbb{Q}),$$

$$t_2 = b_{n-1}\alpha^{n-1} + \cdots + b_1\alpha + a_0 \quad (b_{n-1}, \ldots, b_1, b_0 \in \mathbb{Q}).$$

Let $p(x)$ be the defining monic polynomial of α. Then $\deg(p(x)) = n$. By the help of the Euclidean division we can find polynomials $q(x), r(x) \in \mathbb{Q}[x]$ such that $\deg(r) < n$ and

$$t_1 t_2 = q(\alpha)p(\alpha) + r(\alpha)$$

holds. Since $p(\alpha) = 0$, this gives

$$t_1 t_2 = r(\alpha).$$

However, $r(\alpha) \in T$, thus we get that T is closed with respect to multiplication.

Let now $t \in T$. Then of course the additive inverse of t is $-t$, and we also have $-t \in T$. Thus T is closed with respect to taking additive inverse.

Finally, let $t \in T$, $t \neq 0$. Then we obviously have

$$t = q(\alpha),$$

where $q(x)$ is a polynomial in $\mathbb{Q}[x]$ of degree at most $n-1$. Then as the defining monic polynomial $p(x)$ of α is irreducible, thus

$$\gcd(p(x), q(x)) = 1.$$

Hence by the help of the Euclidean algorithm in the usual way we can find polynomials $u(x), v(x) \in \mathbb{Q}[x]$ for which

$$u(x)p(x) + v(x)q(x) = 1.$$

However, then

$$u(\alpha)p(\alpha) + v(\alpha)q(\alpha) = v(\alpha)q(\alpha) = 1.$$

Hence

$$\frac{1}{t} = \frac{1}{q(\alpha)} = v(\alpha)$$

follows. Here applying division with remainder for v and p, we may assume that $\deg(v(x)) < n$. Thus we get that T is also closed for taking multiplicative inverse, which proves the theorem. $\square$

Remark 10.2.2 For the sake of completeness we note that as one can easily check, if $\alpha \in \mathbb{C}$ is transcendental, then

$$\mathbb{Q}(\alpha) = \left\{ \frac{f(\alpha)}{g(\alpha)} : f(x), g(x) \in \mathbb{Q}[x] \right\},$$

that is, then $\mathbb{Q}(\alpha)$ is isomorphic to the field of rational fractions.

Now we show some examples for algebraic number fields.

Example 10.2.1 The following statements can be easily checked.

1) $K = \mathbb{Q}$ is an algebraic number field of degree one.

2) $K = \mathbb{Q}(\sqrt{2})$ is an algebraic number field of degree two.

3) In general, any algebraic number field of degree two is of the shape $K = \mathbb{Q}(\sqrt{d})$, where $d \neq 0, 1$ is a square-free integer.

In the following we introduce some important notions in algebraic number fields.

Definition 10.2.3 *Let K be an algebraic number field. Then the ring obtained as the intersection of the ring of all algebraic integers with K, is called the ring of integers of K. Notation: O_K.*

Definition 10.2.4 *Let K be an algebraic number field of degree n, and $\omega_1, \ldots, \omega_n$ be a basis of K over $\mathbb{Q}$. If*

$$a_1\omega_1 + \cdots + a_n\omega_n \quad (a_1, \ldots, a_n \in \mathbb{Q})$$

belongs to O_K if and only if $a_1, \ldots, a_n \in \mathbb{Z}$, then $\omega_1, \ldots, \omega_n$ is called an integral basis of K.

Example 10.2.2 Let $K = \mathbb{Q}$. Then K has precisely two integral basis: $\omega_1 = \pm 1$.

Remark 10.2.3 The following assertions are known to be valid.

1) Every algebraic number field possesses an integral basis.

2) The basis transformation matrix between any two integral bases of an algebraic number field is a unimodular matrix.

3) Let $K = \mathbb{Q}(\sqrt{d})$ be a quadratic algebraic number field. (Here $d \neq 0, 1$ is a square-free integer.) Then an integral basis of K is given by $1, \omega$, where

$$\omega = \begin{cases} \sqrt{d}, & \text{if } d \equiv 2, 3 \pmod 4, \\ \frac{1+\sqrt{d}}{2}, & \text{if } d \equiv 1 \pmod 4. \end{cases}$$

10.3 Imaginary quadratic fields

In this section we study imaginary quadratic fields. Our ultimate purpose is to decide whether Euclidean division is available in the ring of integers of such fields K, or in other words, in these cases O_K is a Euclidean ring or not?

First we introduce a norm function on the number fields to be studied.

Definition 10.3.1 *Let $d < 0$ be a square-free integer, $K = \mathbb{Q}(\sqrt{d})$. Then by the norm of an $\alpha \in K$ we mean the number*

$$N(\alpha) = \alpha \cdot \overline{\alpha},$$

where $\overline{\alpha}$ is the complex conjugate of α.

Remark 10.3.1 It is well-known that for any $\alpha \in \mathbb{C}$ we have

$$\alpha \cdot \overline{\alpha} = |\alpha|^2.$$

Thus $N(\alpha)$ is a non-negative real number, moreover, in case of $\alpha \neq 0$ it is a positive real number. We shall use these observations later on without any further mentioning.

In what follows, we discuss some important properties of the norm function introduced above.

Theorem 10.3.1 *Using the previous notation, the norm function holds the following properties.*

1) For any $\alpha \in K$ we have $N(\alpha) \in \mathbb{Q}$, and if $\alpha \in O_K$, then $N(\alpha) \in \mathbb{Z}$.

2) For any $\alpha, \beta \in K$ we have $N(\alpha\beta) = N(\alpha)N(\beta)$.

3) For any $\alpha, \beta \in O_K$, if $\alpha \mid \beta$ (in O_K), then $N(\alpha) \mid N(\beta)$ (in $\mathbb{Z}$).

4) Let $\varepsilon \in O_K$. Then ε is a unit if and only if $N(\varepsilon) = 1$.

5) Let $\alpha \in O_K$. If $N(\alpha)$ is a prime (in $\mathbb{Z}$), then α is irreducible (in O_K).

Proof. We prove the statements separately.

1) If α is rational, then the statement is trivial. Otherwise, observe that the defining monic polynomial of α is just

$$f(x) = x^2 + (\alpha + \overline{\alpha})x + \alpha\overline{\alpha}.$$

This can be seen for example by writing α in the form

$$\alpha = a + b\sqrt{d},$$

where $a, b \in \mathbb{Q}$ - and hence

$$\overline{\alpha} = a - b\sqrt{d}.$$

Thus

$$\alpha\overline{\alpha} = N(\alpha) \in \mathbb{Q}.$$

If $\alpha \in O_K$, then as α is an algebraic integer, we get $f(x) \in \mathbb{Z}[x]$, whence $N(\alpha) \in \mathbb{Z}$ is also valid.

2) Let $\alpha, \beta \in K$. The assertion follows from the next equalities:

$$N(\alpha\beta) = \alpha\beta\overline{\alpha\beta} = \alpha\overline{\alpha}\beta\overline{\beta} = N(\alpha)N(\beta).$$

3) Let $\alpha, \beta \in O_K$. If $\alpha \mid \beta$ in O_K, then

$$\beta = \alpha\gamma$$

holds with some $\gamma \in O_K$. Thus by property 2) we obtain

$$N(\beta) = N(\alpha)N(\gamma).$$

However, then (since by 1) we know that $N(\alpha), N(\beta), N(\gamma) \in \mathbb{Z}$) we get that $N(\alpha) \mid N(\beta)$ in $\mathbb{Z}$, which was to be proved.

4) Let $\varepsilon \in O_K$. If ε is a unit in O_K, then there exists an $\eta \in O_K$, for which

$$\varepsilon\eta = 1$$

holds. Then, since $N(\varepsilon), N(\eta) \in \mathbb{Z}$, and by 3) we have

$$1 = N(1) = N(\varepsilon\eta) = N(\varepsilon)N(\eta),$$

we obtain that

$$N(\varepsilon) = \pm 1.$$

As $N(\varepsilon) \geq 0$, so

$$N(\varepsilon) = 1.$$

Assume now that

$$N(\varepsilon) = 1.$$

Then

$$N(\varepsilon) = \varepsilon\overline{\varepsilon}.$$

Observe that $\overline{\varepsilon} \in O_K$. Indeed, by

$$\overline{\varepsilon} = \frac{1}{\varepsilon}$$

we have $\overline{\varepsilon} \in K$. On the other hand, since $\overline{\varepsilon}$ is the algebraic conjugate of ε as well, $\overline{\varepsilon}$ is an algebraic integer. Thus by the above assertions we conclude that

$$\varepsilon\overline{\varepsilon} = 1$$

is an equality in O_K, thus ε is a unit in O_K.

5) Let $\alpha \in O_K$ such that $N(\alpha) = p$ is prime. Assume to the contrary that α is not irreducible in O_K. Then

$$\alpha = \beta\gamma$$

holds with some non-unit $\beta, \gamma \in O_K$. However, then on the one hand, 4) implies that

$$N(\beta) > 1, \ N(\gamma) > 1,$$

and on the other hand, 2) gives

$$p = N(\alpha) = N(\beta)N(\gamma).$$

This contradiction proves our claim. □

Our next two theorems show that if K is an imaginary quadratic number field, then it may happen that O_K is a Euclidean ring, but it also may happen that it is not.

Theorem 10.3.2 *The ring of the integers of the Gaussian number field, i.e. of $K = \mathbb{Q}(\sqrt{-1})$, is a Euclidean ring.*

Proof. As usual, write O_K for the ring of integers of the Gaussian field $K = \mathbb{Q}(\sqrt{-1})$. Then

$$O_K = \{a + bi : a, b \in \mathbb{Z}\}.$$

Observe that geometrically O_K forms a very nice structure: the elements of O_K are the elements of the (unit) square lattice on the complex plain. Let $\alpha, \beta \in O_K$, $\beta \neq 0$. Let γ be a Gauss-integer (lattice point) on the complex plain whose distance to the number α/β is minimal. Clearly, then we have

$$\left|\frac{\alpha}{\beta} - \gamma\right| \leq \frac{\sqrt{2}}{2} < 1.$$

This immediately gives

$$|\alpha - \beta\gamma| < |\beta|,$$

whence

$$N(\alpha - \beta\gamma) < N(\beta)$$

follows. Let

$$\delta = \alpha - \beta\gamma.$$

Then by the above assertions on the one hand we obtain

$$N(\delta) = N(\alpha - \beta\gamma) < N(\beta),$$

and on the other hand $\alpha, \beta, \gamma \in O_K$ implies $\delta \in O_K$ - since O_K is a ring. However, then we can write

$$\alpha = \beta\gamma + \delta,$$

and here

$$N(\delta) < N(\beta)$$

holds. Thus we proved that O_K is a Euclidean ring indeed, with respect to the norm N. $\square$

Remark 10.3.2 It can be proved in a similar way that the ring of the integers of the Eulerian number field, that is of $K = \mathbb{Q}(\sqrt{-3})$, is a Euclidean ring, too.

Theorem 10.3.3 *Let $K = \mathbb{Q}(\sqrt{-6})$. Then O_K is not a Euclidean ring.*

Proof. We prove by a simple example that not every irreducible element is a prime in O_K. As it is well-known, it proves our claim.

Let

$$\alpha = 2 + \sqrt{-6}.$$

One can easily see that α is irreducible in O_K. Indeed, by

$$N(\alpha) = 10$$

we have that if $\beta \in O_K$ is not a unit and $\beta \mid \alpha$ holds in O_K, then Theorem 10.3.1 implies that

$$N(\beta) = 2 \quad \text{or} \quad N(\beta) = 5$$

holds. However, none of them can be valid, since writing

$$\beta = a + b\sqrt{-6},$$

none of the equations

$$a^2 + 6b^2 = 2, \quad a^2 + 6b^2 = 5$$

is solvable in integers a, b. Thus α is irreducible in O_K, indeed. On the other hand, α is not a prime in O_K. Indeed, consider the identity

$$(2 + \sqrt{-6}) \cdot (2 - \sqrt{-6}) = 10 = 2 \cdot 5$$

in O_K. From this we see that

$$\alpha \mid 2 \cdot 5$$

in O_K. At the same time,

$$N(2) = 4 \quad \text{and} \quad N(5) = 25$$

in view of $N(\alpha) = 10$ implies that

$$\alpha \nmid 2 \quad \text{and} \quad \alpha \nmid 5$$

in O_K. That is, α divides a product in O_K, however, it does not divide any of the terms. Therefore α is not a prime, which proves our claim. $\square$

Remark 10.3.3 We note that in fact O_K is only 'rarely' Euclidean among imaginary quadratic number fields K.

10.4 Exercises

Exercise 10.4.1 Let α be an algebraic number, ϑ be a transcendental number. Prove that then

- $\alpha + \vartheta$ is transcendental,
- $\alpha\vartheta$ is algebraic if and only if $\alpha = 0$.

Exercise 10.4.2 Show that the sum and the product of two transcendental numbers can be both algebraic and transcendental!

Exercise 10.4.3 Prove that $\pi + \sqrt{2}i$ is transcendental!

Exercise 10.4.4 Let α and β be algebraic numbers. Prove that if $\alpha + \beta$ is rational, then

$$\deg(\alpha) = \deg(\beta)$$

holds!

Exercise 10.4.5 Let α be an algebraic number. Prove that then for any $m \geq 2$ the number $\sqrt[m]{\alpha}$ is algebraic, and

$$\deg(\sqrt[m]{\alpha}) \leq m \deg(\alpha)$$

holds!

Prove also that if α is an algebraic integer, then $\sqrt[m]{\alpha}$ is also an algebraic integer!

Exercise 10.4.6 Let $\alpha \neq 0$ be an algebraic integer. When will $1/\alpha$ be an algebraic integer, too?

Exercise 10.4.7 Prove that $\cos 20^\circ$ is a cubic algebraic number!

Exercise 10.4.8 Prove that $1/\sin 22,5^\circ$ is an algebraic integer, and determine its degree!

Exercise 10.4.9 Let

$$f(x) = x^6 - x^5 + 2x^4 - x^3 + 2x^2 - 4x - 2.$$

We know that $\sqrt[3]{2}$ is a root of f. Prove that every root of $f(x)$ is an algebraic integer, of degree at most three!

Exercise 10.4.10 Give the defining monic polynomials of the following algebraic numbers:

- $\sqrt{2}$
- i
- -1
- $\sqrt{2}+1$
- $1-2i$
- $\frac{1+\sqrt{5}}{2}$
- $\sqrt[3]{2}-1$

Exercise 10.4.11 Let $K = \mathbb{Q}(\sqrt{2}+1)$ and $L = \mathbb{Q}(\sqrt{8})$. Prove that $K = L$.

Exercise 10.4.12 Let K and L be quadratic algebraic number fields. Prove that then

$$K = L \quad \text{or} \quad K \cap L = \mathbb{Q}$$

holds!

Exercise 10.4.13 Let α be a root of the polynomial $f(x) = x^3 - x - 1$. Prove that α is a cubic algebraic integer!

Further, let $K = \mathbb{Q}(\alpha)$ and $\beta \in K$,

$$\beta = \alpha^6 - \alpha^5 + \alpha^3 + \alpha^2 - \alpha - 2.$$

Represent β in the form

$$\beta = a\alpha^2 + b\alpha + c \quad (a, b, c \in \mathbb{Q}).$$

Exercise 10.4.14 Let $K = \mathbb{Q}(\sqrt[3]{2})$, $\alpha, \beta, \gamma \in K$,

$$\alpha = 1 - \sqrt[3]{2}, \quad \beta = 1 + \sqrt[3]{2} + \sqrt[3]{4}, \quad \gamma = -\sqrt[3]{4} + \sqrt[3]{16}.$$

Show that α, β, γ form a basis of K over $\mathbb{Q}$!

Exercise 10.4.15 Let $K = \mathbb{Q}(\sqrt{-1})$ be the Gaussian number field, O_K be its ring of integers. Decide the validity of the following divisibility relations in O_K:

- $1 + i \mid 9 - 13i$
- $1 + 3i \mid 24 + 21i$
- $2 + 3i \mid 10 - 47i$

Exercise 10.4.16 Let $K = \mathbb{Q}(\sqrt{-1})$ be the Gaussian number field, O_K its ring of integers. Find the greatest common divisors of the Gauss integers α, β, and express them as a linear combination of α, β:

- $\alpha = 6 + 6i$, $\beta = 5 + 3i$
- $\alpha = 3 + 11i$, $\beta = 4 + 7i$

Exercise 10.4.17 Let $K = \mathbb{Q}(\sqrt{-1})$ be the Gaussian number field, O_K its ring of integers. Give the prime factorization of α in the ring of the Gauss integers:

- $\alpha = 2$
- $\alpha = 4 + 6i$

- $\alpha = 7 + 2i$

Exercise 10.4.18 Let $K = \mathbb{Q}(\sqrt{-3})$ be the Eulerian number field, $\alpha, \beta \in O_K$,

$$\alpha = \frac{17}{2} - \frac{13}{2}\sqrt{-3}, \quad \beta = 23.$$

Find the greatest common divisor of α and β.

Exercise 10.4.19 Let $K = \mathbb{Q}(\sqrt{-3})$ be the Eulerian number field, $\alpha \in O_K$,

$$\alpha = 12 - \sqrt{-3}.$$

Express α as a product of primes in O_K.

Chapter 11

Diophantine approximation

In this chapter we study the approximation of real numbers with rational numbers.

11.1 Basic notions and assertions

We start with introducing the basic concept of the topic.

Definition 11.1.1 *Let $\alpha \in \mathbb{R}$. If for some positive real number κ we can find a constant $c(\alpha)$ depending only on α and infinitely many numbers $x_n, y_n \in \mathbb{Z}$ $(n \in \mathbb{N})$ with $\gcd(x_n, y_n) = 1$ and $x_n > 0$ such that*

$$\left|\alpha - \frac{y_n}{x_n}\right| < \frac{c(\alpha)}{x_n^{\kappa}} \quad (n \in \mathbb{N})$$

holds, then we say that α can be approximated in order κ. The rational numbers x_n/y_n are called approximating fractions.

Remark 11.1.1 Observe that if $\kappa_1 > \kappa_2 > 0$ and α can be approximated in order κ_1, then clearly, α can be approximated in order κ_2, as well. So the interesting question in case of a given real number α is that in how large order can the number α be approximated.

We also mention that the essence of the topic (as we have already mentioned) is the problem of approximation of real numbers by rational numbers. As $\mathbb{Q}$ is a dense subset of $\mathbb{R}$, hence it is clear that we must find a way to measure the 'quality' of approximation. This is done by the help of the order of

the approximation: it is the exponent κ, on which raising the denominator, we still have infinitely many rationals satisfying the required inequality. The role of the value of $c(\alpha)$ is much less significant.

Our first theorem is in fact an immediate consequence of the later ones. However, the proof gives a viewpoint which is helpful in the topic, so its discussion is useful.

Theorem 11.1.1 *Every real number can be approximated in order one.*

Proof. We follow the notation of the definition. Let $\alpha \in \mathbb{R}$ be arbitrary, and let $c(\alpha) = 2$. Let $x_n = p_n$ be the n-th prime ($n \in \mathbb{N}$). For a given n consider the rational numbers z/x_n, where $z \in \mathbb{Z}$. Obviously, α is contained in an interval of the shape $[z/x_n, (z+1)/x_n]$, with some $z \in \mathbb{Z}$. It is also clear that one of the numbers z and $z+1$ is coprime to x_n; denote it by y_n. (If both values are coprime to x_n, we may choose any of them.) Hence of course

$$\left|\alpha - \frac{y_n}{x_n}\right| \leq \frac{1}{x_n} < \frac{2}{x_n} = \frac{c(\alpha)}{x_n} \quad (n \in \mathbb{N}).$$

By the choice of the numbers x_n the fractions y_n/x_n are distinct, which proves the statement. □

The next result provides a simple, but very important assertion.

Theorem 11.1.2 *Let $\kappa > 0$ and suppose that $\alpha \in \mathbb{R}$ can be approximated in order κ. Then the sequence of the denominators of the approximating fractions tends to infinity.*

Proof. We follow the notation of the definition. Let $\alpha \in \mathbb{R}$, and let $c(\alpha)$ be arbitrary, but fixed. Assume to the contrary that the sequence x_n of the denominators of the approximating fractions does not tend to infinity. Then there exists a value x_m which occurs infinitely often. However, then by our assumptions for this value x_m we can find infinitely many distinct numerators y_n, that is

$$\left|\alpha - \frac{y_n}{x_m}\right| < \frac{c(\alpha)}{x_m^\kappa} \leq c(\alpha)$$

holds for infinitely many distinct y_n. But this is not possible: there cannot exist infinitely many rational numbers with the same denominator, all closer to some real number α than a fixed distance. This contradiction proves our claim. □

11.2 Approximation of rational, irrational and algebraic numbers

In this section we investigate the approximation properties of different types of real numbers. We start with studying rational numbers. It will turn out that the first order approximation proved in the previous section is in fact the final stop for these numbers.

Theorem 11.2.1 *The rational numbers can be approximated precisely in first order.*

Proof. We have seen already that every real number, hence every rational number can be approximated in order one. However, for this property - to widen our insight - we give a different proof.

Let $\alpha \in \mathbb{Q}$, $\alpha = a/b$, where a, b are coprime integers with $b > 0$. Let

$$c(\alpha) = \frac{2}{b}.$$

Consider the equations

$$ax - by = \pm 1.$$

These, by $\gcd(a, b) = 1$ have infinitely many solutions. Let x_n, y_n be a solution of one of them with $x_n > 0$. (If here $x_n < 0$ would hold, then consider the numbers $-x_n, -y_n$.) Then we clearly also have $\gcd(x_n, y_n) = 1$, further,

$$\left|\frac{a}{b} - \frac{y_n}{x_n}\right| = \frac{1}{bx_n} < \frac{c(\alpha)}{x_n}.$$

This shows that α can be approximated in order one.

Assume now to the contrary that α can be approximated in some order $\kappa > 1$. Then, using the previous notation we have

$$\left|\frac{a}{b} - \frac{y_n}{x_n}\right| < \frac{c(\alpha)}{x_n^\kappa}$$

with some value $c(\alpha)$, for infinitely many approximating fractions y_n/x_n. From this we get

$$|ax_n - by_n| < \frac{c(\alpha)}{x_n^{\kappa-1}}.$$

Using Theorem 11.1.2 from the previous section, we can see that the right hand side of the above expression tends to infinity (as n tends to infinity). Hence we obtain

$$\lim_{n\to\infty} |ax_n - by_n| = 0.$$

However, $|ax_n - by_n|$ is an integer! Thus we conclude that there exists an integer n_0, such that for $n > n_0$ we have

$$ax_n - by_n = 0,$$

that is

$$\frac{a}{b} = \frac{y_n}{x_n}.$$

However, this is impossible, since the approximating fractions must be different. This contradiction proves our theorem. □

Remark 11.2.1 In the formulation of the theorem, and also later on, we use the usual phrases of the field. Namely, the phrase 'can be approximated precisely in order one' in fact means that the rational numbers can be approximated in order one, however, they cannot be approximated in any order greater than one. On the other hand, they can obviously approximated in any order smaller than one (as we already mentioned in more general circumstances).

Now we formulate a famous theorem of Dirichlet, saying that the irrational numbers can be approximated in order at least two. To prove this theorem, we shall need the following lemma, also due to Dirichlet.

Lemma 11.2.1 *Let α be a real number, and Q be an integer with $Q > 1$. Then there exist integers p, q such that*

$$1 \leq q \leq Q \quad \textit{and} \quad |q\alpha - p| \leq \frac{1}{Q}$$

holds.

Proof. We give two proofs of the statement, relying on different principles. The first one is the original argument of Dirichlet, based upon the pigeonhole principle. The second proof - maybe surprisingly - uses the theorem of Minkowski.

First proof. For an arbitrary $\xi \in \mathbb{R}$ let $\lfloor \xi \rfloor$ be the integer part of ξ, and $\{\xi\}$ be the fractional part of ξ, that is

$$\lfloor \xi \rfloor = \max_{z \in \mathbb{Z},\ z \le \xi} z \quad \text{and} \quad \{\xi\} = \xi - \lfloor \xi \rfloor.$$

We clearly have

$$0 \le \{\xi\} < 1.$$

Consider the numbers

$$0,\ \{\alpha\},\ \{2\alpha\},\ \ldots,\ \{(Q-1)\alpha\},\ 1.$$

Obviously, all these $Q+1$ numbers belong to the interval $[0,1]$. Thus necessarily there exist two of them whose distance is at most $1/Q$. Clearly, these two numbers cannot be 0 and 1. If one of these numbers is 0, the other one is $\{t\alpha\}$ $(1 \le t \le Q-1)$, then by

$$|\{t\alpha\} - 0| \le \frac{1}{Q}$$

using the notation

$$q = t \quad \text{and} \quad p = \lfloor t\alpha \rfloor$$

we obtain

$$|q\alpha - p| \le \frac{1}{Q}.$$

The situation is similar if one of the numbers is $\{t\alpha\}$ $(1 \le t \le Q-1)$, and the other one is 1. Indeed, then using

$$|1 - \{t\alpha\}| \le \frac{1}{Q}$$

with the notation

$$q = t \quad \text{and} \quad p = \lfloor t\alpha \rfloor + 1$$

we obtain

$$|q\alpha - p| \le \frac{1}{Q}.$$

Finally, assume that the distance of $\{t\alpha\}$ and $\{k\alpha\}$ $(1 \le t < k \le Q-1)$ is smaller than $1/Q$. Then by

$$|\{k\alpha\} - \{t\alpha\}| \le \frac{1}{Q}$$

with the notation

$$q = k - t \quad \text{and} \quad p = \lfloor k\alpha \rfloor - \lfloor t\alpha \rfloor$$

we get

$$|q\alpha - p| \le \frac{1}{Q}.$$

That is, our claim holds in any case.

Second proof. Consider the lattice $\Lambda = \mathbb{Z}^2$ in $\mathbb{R}^2$. Then of course Λ is a full lattice, with lattice determinant $\det(\Lambda) = 1$. Further, let

$$S = \left\{ (q,p) : q,p \in \mathbb{R},\ |q| \le Q,\ |q\alpha - p| \le \frac{1}{Q} \right\}.$$

Observe that then S is a closed parallelogram (containing its sides), which is symmetric with respect to the origin, that is, it is a convex, centralsymmetric set. It is also easy to check that the area of S is

$$\mathrm{V}(S) = 2Q \cdot \frac{2}{Q} = 4 = 2^2 \cdot 1 = 2^2 \det(\Lambda).$$

Thus by part ii) of the Minkowski theorem we get that there exist $q,p \in \mathbb{Z}$, $(q,p) \neq (0,0)$ such that

$$|q| \le Q \quad \text{and} \quad |q\alpha - p| \le \frac{1}{Q}.$$

Observe that if here we would have $q = 0$, then by $Q > 1$ we would also have $p = 0$, which by $(q,p) \neq (0,0)$ is impossible. Hence necessarily $q \neq 0$. However, then by the centralsymmetry of S, we may also assume that $q \ge 1$. Indeed, if $q < 1$ then $q < 0$, and in place of the point (q,p) taking the point $(-q,-p)$ the above inequalities remain valid. Thus altogether

$$q,p \in \mathbb{Z}, \quad 1 \le q \le Q \quad \text{and} \quad |q\alpha - p| \le \frac{1}{Q}$$

hold, which proves the statement. □

The next theorem is also due to Dirichlet. This shows that the irrational numbers can be approximated in order two, and this approximation is uniform in the sense that for any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ one can take $c(\alpha) = 1$.

Theorem 11.2.2 (Dirichlet theorem) *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Then there exist infinitely many different $x_n, y_n \in \mathbb{Z}$ with $\gcd(x_n, y_n) = 1$ and $x_n > 0$, such that*

$$\left|\alpha - \frac{y_n}{x_n}\right| < \frac{1}{x_n^2}.$$

Proof. By Lemma 11.2.1 for every $n > 1$ we can find integers x_n, y_n for which

$$1 \le x_n \le n \quad \text{and} \quad |\alpha x_n - y_n| \le \frac{1}{n}$$

hold. Here we may further assume that $\gcd(x_n, y_n) = 1$. Indeed, if $\gcd(x_n, y_n) = d > 1$ would hold, then writing $x_n' = x_n/d$, $y_n' = y_n/d$ we would have

$$1 \le x_n' \le n \quad \text{and} \quad |\alpha x_n' - y_n'| \le \frac{1}{nd} \le \frac{1}{n}$$

as well. Thus in place of the pair (x_n, y_n) we could take the pair (x_n', y_n').

By the above assertions we get

$$\left|\alpha - \frac{y_n}{x_n}\right| \le \frac{1}{nx_n} \le \frac{1}{x_n^2}.$$

Since α is irrational, necessarily

$$\left|\alpha - \frac{y_n}{x_n}\right| < \frac{1}{x_n^2}$$

holds. That is, the integers x_n, y_n satisfy the requirements of the theorem. So it is only left to show that there are infinitely many such numbers. This follows from the inequality

$$\left|\alpha - \frac{y_n}{x_n}\right| < \frac{1}{n},$$

since a fixed pair (x_{n_0}, y_{n_0}) can clearly satisfy it only for finitely many values of n. This completes the proof. □

For the sake of clarity we formulate the following statement, which is a trivial consequence of the previous theorem.

Corollary 11.2.1 *The irrational numbers can be approximated in order two.*

Finally, we consider approximations to algebraic numbers. In the following we discuss a famous theorem of Liouville, which shows that algebraic numbers can be approximated only in bounded orders.

Theorem 11.2.3 (Liouville theorem) *An algebraic number of degree k cannot be approximated in order larger than k.*

Proof. Let α be an algebraic number of degree k. Since rational numbers can be approximated precisely in order one, we may assume that $k \geq 2$. Let $g(z) \in \mathbb{Q}[z]$ be the defining monic polynomial of α. Then there exists a positive integer a_0 such that $f(z) = a_0 g(z)$ is a primitive polynomial with integer coefficients. (In fact, a_0 is just the least common multiple of the denominators of the coefficients of $g(z)$.) Let

$$f(z) = a_0 z^k + a_1 z^{k-1} + \cdots + a_{k-1} z + a_k = a_0 (z - \alpha_1) \dots (z - \alpha_k),$$

where $\alpha_1, \dots, \alpha_k$ are the algebraic conjugates of α; here we may choose $\alpha = \alpha_1$. Since $f(z)$ is irreducible over $\mathbb{Q}$ (and $k \geq 2$), thus $\alpha_i \notin \mathbb{Q}$ $(i = 1, \dots, k)$. We also obtain that the roots $\alpha_1, \dots, \alpha_k$ are pairwise distinct. Assume to the contrary that α can be approximated in order $k + \varepsilon$, where ε is a positive real number. Then with some value $c(\alpha)$ and infinitely many approximating fractions y_n/x_n

$$\left| \alpha - \frac{y_n}{x_n} \right| < \frac{c(\alpha)}{x_n^{k+\varepsilon}}. \tag{11.1}$$

Hence, as y_n/x_n cannot be a root of f, and as

$$\left| x_n^k f\left(\frac{y_n}{x_n} \right) \right|$$

is an integer, we obtain that

$$1 \leq \left| x_n^k f\left(\frac{y_n}{x_n} \right) \right| = x_n^k a_0 \left| \alpha - \frac{y_n}{x_n} \right| \prod_{i=2}^{k} \left| \alpha_i - \frac{y_n}{x_n} \right|. \tag{11.2}$$

As

$$\left| \alpha_i - \frac{y_n}{x_n} \right| \leq |\alpha - \alpha_i| + \left| \alpha - \frac{y_n}{x_n} \right| < |\alpha - \alpha_i| + c(\alpha) \quad (i = 2, \dots, k),$$

thus there exists a value $c^*(\alpha)$ depending only on α for which, based upon (11.2) and (11.1),

$$1 \leq c^*(\alpha) x_n^k \left| \alpha - \frac{y_n}{x_n} \right| < \frac{c^*(\alpha) c(\alpha)}{x_n^{\varepsilon}}$$

holds. As Lemma 11.1.2 gives

$$\lim_{n \to \infty} x_n \to \infty,$$

we get a contradiction. This proves the theorem. □

By the help of the above theorem Liouville could construct transcendental numbers. This is shown by the following corollary.

Corollary 11.2.2 *The number*

$$\alpha = \sum_{i=0}^{\infty} \frac{1}{2^{i!}}$$

is transcendental.

Proof. We prove that α can be approximated in arbitrary order - this in view of the previous theorem of Liouville implies our claim. It is clear that α is well-defined, that is, the series defining α is convergent, since it is majorated by the convergent geometric series

$$\sum_{i=0}^{\infty} \frac{1}{2^i}.$$

For an arbitrary $n \in \mathbb{N}$ let $x_n = 2^{n!}$ and

$$y_n = 2^{n!} \sum_{i=0}^{n} \frac{1}{2^{i!}}.$$

Observe that then y_n is an odd integer for any $n \in \mathbb{N}$, hence $\gcd(x_n, y_n) = 1$. Further, for any $n \in \mathbb{N}$ we have

$$\left| \alpha - \frac{y_n}{x_n} \right| = \sum_{i=n+1}^{\infty} \frac{1}{2^{i!}} < \sum_{i=(n+1)!}^{\infty} \frac{1}{2^i} =$$

$$= \frac{1}{2^{(n+1)!}} \left(1 + \frac{1}{2} + \frac{1}{4} + \dots \right) = \frac{2}{(2^{n!})^{n+1}} = \frac{2}{x_n^{n+1}}.$$

However, then for any $\kappa \geq 1$ we get that α can be approximated in order κ, since the fractions y_n/x_n $(n \geq \kappa - 1)$ are approximating fractions. That is, α can be approximated in any order indeed, and our statement follows. $\square$

Remark 11.2.2 In spite of that 'almost all' real numbers are transcendental (as the algebraic numbers form a set of Lebesgue measure zero), interestingly it is not that easy to prove from a concrete transcendental number that it is indeed transcendental. In fact, the above so-called Liouville-type number is the first example of a transcendental number, for which we proved that it is transcendental indeed. The real numbers 'similar' to α, that is, which can be approximated in arbitrary order (which hence are transcendental) are usually called Liouville-numbers.

At the end of the section, we give without a proof a famous theorem of Roth, which gives the best possible description of approximations of irrational numbers. (We note that for this result Roth received the Fields-medal.)

Theorem 11.2.4 (Roth theorem) *An algebraic number cannot be approximated in order larger than two.*

11.3 Exercises

Exercise 11.3.1 For the following α find all fractions p/q, for which

$$1 \leq q \leq 10$$

and

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$$

holds:

- $\alpha = \pi$
- $\alpha = e$
- $\alpha = \sqrt{2}$

Exercise 11.3.2 Let α be an arbitrary real number. Prove that there exist infinitely many $k \in \mathbb{N}$ and $p \in \mathbb{Z}$ such that

$$\left|\alpha - \frac{p}{2^k}\right| \leq \frac{1}{3} \cdot \frac{1}{2^k}$$

holds!

Exercise 11.3.3 Prove that there exists a real number α such that for any $k \in \mathbb{N}$ and $p \in \mathbb{Z}$ we have

$$\left|\alpha - \frac{p}{3^k}\right| \geq \frac{1}{2} \cdot \frac{1}{3^k}.$$

Exercise 11.3.4 Prove that the sequence

$$a_n = \sin(\sqrt{2}+1)^n \pi \quad (n = 1, 2, 3, \dots)$$

is convergent and its limit is 0.

Exercise 11.3.5 Prove that for any $\varepsilon > 0$ there exist infinitely many $n \in \mathbb{N}$ such that

$$|\sin n| < \varepsilon$$

holds!

Exercise 11.3.6 Prove that

$$\sum_{n=0}^{\infty} \frac{1}{3^{n!}}$$

is a transcendental number!

Chapter 12

Continued fractions and their applications

In this chapter we study the continued fraction representations of rational and irrational numbers, and certain applications of continued fractions.

12.1 Finite continued fractions

In this section we discuss finite continued fractions.

Definition 12.1.1 *Let a_0 be a real number, and $a_1, \ldots, a_n$ positive real numbers. Then the expression*

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}}}$$

is called a finite continued fraction, the numbers $a_0, a_1, \ldots, a_n$ are called the digits of the continued fraction. If $a_0, a_1, \ldots, a_n \in \mathbb{Z}$, then we say that the continued fraction is simple. Since the above formula is rather complicated, in what follows we shall use the notation

$$[a_0; a_1, \ldots, a_n]$$

for continued fractions instead.

Example 12.1.1 Consider the number 62/23. Then we can write

$$\frac{62}{23} = 2 + \frac{16}{23} = 2 + \cfrac{1}{1 + \cfrac{7}{16}} = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{2}{7}}} = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{2}}}}$$

that is,

$$\frac{62}{23} = [2; 1, 2, 3, 2].$$

The next theorem shows that there is an (essentially) one-to-one correspondence between rational numbers and finite, simple continued fractions.

Theorem 12.1.1 *Every finite, simple continued fraction represents a rational number. If* $[a_0; a_1, \ldots, a_n]$ *is a finite, simple continued fraction and* $n = 0$ *or* $a_n > 1$*, then*

$$[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_n - 1, 1]$$

holds. Apart from this identity, the continued fraction representation of any rational number is unique.

Proof. If $[a_0; a_1, \ldots, a_n]$ is a finite, simple continued fraction, then by the definition we immediately see that its value is a rational number. If $n = 0$ or $a_n > 1$, then the identity

$$[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_n - 1, 1]$$

by $a_n = (a_n - 1) + 1$ trivially holds.

Now we show that every rational number can be represented as a finite, simple continued fraction. For this, let $q \in \mathbb{Q}$, $q = a/b$, where a, b are coprime

integers with $b > 0$. Execute the Euclidean algorithm with the numbers a, b. Then we obtain

$$\begin{aligned} r_0 &= q_1 r_1 + r_2 \\ r_1 &= q_2 r_2 + r_3 \\ &\vdots \\ r_{n-2} &= q_{n-1} r_{n-1} + r_n \\ r_{n-1} &= q_n r_n \end{aligned}$$

where $r_0 = a$, $r_1 = b$. From this we get

$$q = \frac{a}{b} = \frac{r_0}{r_1} = q_1 + \frac{r_2}{r_1} = q_1 + \cfrac{1}{q_2 + \cfrac{r_2}{r_3}} = \cdots = q_1 + \cfrac{1}{q_2 + \cfrac{1}{\cdots + \cfrac{1}{q_{n-1} + \cfrac{1}{q_n}}}}.$$

Hence we conclude that

$$q = [q_1; q_2, \ldots, q_n],$$

that is, q can be represented as a finite, simple continued fraction indeed.

Thus we only need to prove that the representation of rational numbers as finite, simple continued fractions (apart from the above mentioned identity) is unique. For this purpose, let $q \in \mathbb{Q}$ and let

$$q = [a_0; a_1, \ldots, a_k] = [b_0; b_1, \ldots, b_\ell]$$

be two representations of q as finite, simple continued fractions. We may clearly assume that $k \leq \ell$. We prove the statement by induction on k.

Take first $k = 0$. Then of course

$$q = a_0.$$

Hence necessarily $q \in \mathbb{Z}$. Observe that

$$q = b_0 + \frac{1}{[b_1; b_2, \ldots, b_\ell]}$$

and here

$$[b_1; b_2, \ldots, b_\ell] > 1 \quad \text{or} \quad [b_1; b_2, \ldots, b_\ell] = 1,$$

but in the latter case we have $\ell = 1$ and $b_1 = 1$. Hence we obtain that either $\ell = 0$, when

$$q = b_0,$$

or $\ell = 1$ and

$$q = b_0 + 1.$$

Therefore in this case our claim follows, since the two representations of q are given by either

$$[a_0] = q = [b_0] \quad \text{and} \quad a_0 = b_0$$

or

$$[a_0] = q = [b_0; 1] \quad \text{and} \quad a_0 = b_0 + 1.$$

Assume now that the statement is valid for some $k = m$, and consider that with $k = m + 1$. Observe that we may suppose that

$$[a_1; a_2, \ldots, a_k] > 1$$

and

$$[b_1; b_2, \ldots, b_\ell] > 1.$$

Indeed, otherwise $k = 1$, $a_1 = 1$ and $\ell = 1$, $b_1 = 1$ would hold, respectively. However, then in place of the representation

$$q = [a_0; a_1]$$

or

$$q = [b_0; b_1]$$

we could write

$$q = [a_0]$$

or

$$q = [b_0],$$

respectively - and this possibility (by symmetry) is included in the already discussed case $k = 0$. Thus the representations

$$q = a_0 + \frac{1}{[a_1; a_2, \ldots, a_k]}$$

and

$$q = b_0 + \frac{1}{[b_1; b_2, \ldots, b_\ell]}$$

immediately give

$$a_0 = \lfloor q \rfloor$$

and

$$b_0 = \lfloor q \rfloor,$$

whence

$$a_0 = b_0,$$

furthermore,

$$q - a_0 = [a_1; a_2, \ldots, a_k]$$

and

$$q - b_0 = [b_1; b_2, \ldots, b_\ell]$$

follow. Hence from the induction hypothesis we get $\ell = k$ and

$$a_i = b_i \quad (i = 1, \ldots, k),$$

or $\ell = k + 1$ and

$$a_i = b_i \ (i = 1, \ldots, k) \quad \text{and} \quad b_\ell = 1.$$

This proves the theorem. □

Remark 12.1.1 The above theorem and its proof shows that the (essentially unique) representation of a rational number a/b as a finite, simple continued fraction is given by the quotients $q_1, \ldots, q_n$, appearing during the Euclidean algorithm executed with a, b.

12.2 Infinite continued fractions

In this section we study irrational numbers and infinite continued fractions attached to them.

Definition 12.2.1 *Let α be an irrational number. Define the following numbers inductively:*

$$a_0 = \lfloor \alpha \rfloor, \ \alpha_1 = \frac{1}{\alpha - a_0}, \ a_1 = \lfloor \alpha_1 \rfloor, \ \alpha_2 = \frac{1}{\alpha_1 - a_1}, \ \ldots,$$
$$a_n = \lfloor \alpha_n \rfloor, \ \alpha_{n+1} = \frac{1}{\alpha_n - a_n}, \ \ldots \ .$$

Then

$$[a_0; a_1, \dots, a_n, \dots]$$

is called the infinite continued fraction belonging to α.

Remark 12.2.1 Since α is irrational, thus one can easily see that the numbers $\alpha_1, \alpha_2, \dots$ are also irrational. Thus the numbers $\alpha_i - a_i$ are non-zero, so the above definition works for all n.

In this section our main purpose is to show that the above continued fraction determines the irrational number α uniquely. For this, the following trivial observation will be of great help.

Lemma 12.2.1 *Let* α *be an irrational number. Then by the notation of Definition 12.2.1*

$$\alpha = [a_0; a_1, \dots, a_{n-1}, \alpha_n]$$

holds for all $n \geq 0$. *(Here* $\alpha_0 = \alpha$.*)*

Proof. The statement follows from the next sequence of equalities, which trivially hold by the definition:

$$\alpha = \alpha_0 = a_0 + \frac{1}{\alpha_1} = [a_0; \alpha_1] = \left[a_0; a_1 + \frac{1}{\alpha_2}\right] = [a_0; a_1, \alpha_2] =$$
$$= \dots = \left[a_0; a_1, \dots, a_{n-2}, a_{n-1} + \frac{1}{\alpha_n}\right] = [a_0; a_1, \dots, a_{n-2}, a_{n-1}, \alpha_n].$$

□

Remark 12.2.2 In the above lemma the continued fraction

$$[a_0; a_1, \dots, a_{n-1}, \alpha_n]$$

is clearly not simple, since α_n is an irrational number (not an integer).

The following sequences will play rather important roles later on.

Definition 12.2.2 *Let* $[a_0; a_1, a_2, \dots]$ *be an infinite continued fraction belonging to an irrational number* α. *Introduce the following sequences* b_n *and* c_n $(n \geq -2)$*:*

$$b_{-2} = 0,\ b_{-1} = 1 \quad \textit{and} \quad c_{-2} = 1,\ c_{-1} = 0,$$

and

$$b_n = a_n b_{n-1} + b_{n-2}\ (n \geq 0) \quad \textit{and} \quad c_n = a_n c_{n-1} + c_{n-2}\ (n \geq 0).$$

The next theorem provides some important properties of the sequences b_n and c_n.

Theorem 12.2.1 *Let b_n and c_n be as above. Then the following assertions hold:*

i) $\lim_{n\to\infty} c_n = \infty$,

ii) $b_n c_{n-1} - b_{n-1}c_n = (-1)^{n-1}$ $(n \geq -1)$,

iii) $\gcd(b_n, c_n) = 1$.

Proof. We prove the statements separately.

i) By the definition of c_n, for the first few terms of the sequence we have

$$c_{-2} = 1,\ c_{-1} = 0,\ c_0 = 1,\ c_1 = a_1,\ c_2 = a_1a_2 + 1.$$

In particular,

$$c_n \geq 1 \quad (n \geq 0).$$

Thus

$$c_{n+1} = a_{n+1}c_n + c_{n-1} > c_n \quad (n \geq 1),$$

which proves our claim.

ii) We prove the statement by induction. If $n = -1$, then

$$b_{-1}c_{-2} - b_{-2}c_{-1} = 1 \cdot 1 - 0 \cdot 0 = 1 = (-1)^{-2},$$

so the statement is valid. Assume now that the statement holds for some $n = k \geq -1$, that is

$$b_k c_{k-1} - b_{k-1}c_k = (-1)^{k-1}.$$

Then by the definitions of the sequences b_n and c_n, using the induction hypothesis we get

$$\begin{aligned} b_{k+1}c_k - b_k c_{k+1} = (a_{k+1}b_k + b_{k-1})c_k - b_k(a_{k+1}c_k + c_{k-1}) = \\ = b_{k-1}c_k - b_k c_{k-1} = -(-1)^{k-1} = (-1)^k, \end{aligned}$$

so the assertion is valid also for $n = k + 1$. Hence our claim follows.

iii) Let $n \geq -2$ and

$$d = \gcd(b_n, c_n).$$

Then clearly

$$d \mid b_{n+1}c_n - b_n c_{n+1}.$$

Thus by part ii) we obtain $d = 1$, which proves the statement. □

The next lemma is of technical nature. In fact, it is important for the following Theorem 12.2.2, which reveals the importance of the sequences b_n and c_n.

Lemma 12.2.2 *Let $[a_0; a_1, a_2, \dots]$ be the infinite continued fraction belonging to the irrational number α, and let b_n and c_n be the sequences introduced in Definition 12.2.2. Then for any $n \geq 0$ and positive real number t we have*

$$[a_0; a_1, \dots, a_{n-1}, t] = \frac{tb_{n-1} + b_{n-2}}{tc_{n-1} + c_{n-2}}.$$

Proof. When $n = 0$, we have to check the validity of the assertion

$$[t] = \frac{tb_{n-1} + b_{n-2}}{tc_{n-1} + c_{n-2}},$$

which by

$$b_{-2} = 0,\ b_{-1} = 1,\ c_{-2} = 1,\ c_{-1} = 0$$

trivially holds. In case of $n \geq 1$ the statement is proved by induction.

Let $n = 1$. Then by

$$b_0 = a_0 \quad \text{and} \quad c_0 = 1$$

we get

$$[a_0; t] = a_0 + \frac{1}{t} = \frac{ta_0 + 1}{t} = \frac{tb_0 + b_{-1}}{tc_0 + c_{-1}},$$

which proves our claim. Assume now that the statement is valid for some $n = k \geq 1$, that is

$$[a_0; a_1, \dots, a_{k-1}, t] = \frac{tb_{k-1} + b_{k-2}}{tc_{k-1} + c_{k-2}}$$

for any positive real number t. Then by the induction hypothesis we obtain

$$[a_0; a_1, \dots, a_{k-1}, a_k, t] = \left[a_0; a_1, \dots, a_{k-1}, a_k + \frac{1}{t}\right] =$$

$$= \frac{\left(a_k + \frac{1}{t}\right) b_{k-1} + b_{k-2}}{\left(a_k + \frac{1}{t}\right) c_{k-1} + c_{k-2}} = \frac{t(a_k b_{k-1} + b_{k-2}) + b_{k-1}}{t(a_k c_{k-1} + c_{k-2}) + c_{k-1}} = \frac{tb_k + b_{k-1}}{tc_k + c_{k-1}},$$

so the assertion is valid for $n = k + 1$, as well. Hence our statement follows. $\square$

Our next theorem shows the real importance of the sequences b_n and c_n: they provide the numerators and denominators of the partial continued fractions, as rational numbers, of the infinite continued fraction belonging to the irrational number appearing in their definitions.

Theorem 12.2.2 *Let α be an irrational number, and $[a_0; a_1, a_2, \ldots, a_n, \ldots]$ be the infinite continued fraction belonging to α. Let r_n be the rational number determined by the n-th partial continued fraction, that is*

$$r_n = [a_0; a_1, a_2, \ldots, a_n] \quad (n \geq 0).$$

Then we have

$$r_n = \frac{b_n}{c_n} \quad (n \geq 0),$$

where b_n and c_n are the sequences appearing in Definition 12.2.2. Further, here r_n is given in primitive form, that is, $c_n > 0$ and the fraction cannot be simplified.

Proof. By Lemma 12.2.2 we know that for any $n \geq 0$, by the choice $t = a_n$ we obtain

$$[a_0; a_1, a_2, \ldots, a_n] = \frac{a_n b_{n-1} + b_{n-2}}{a_n c_{n-1} + c_{n-2}} = \frac{b_n}{c_n} = r_n.$$

The fact that for any $n \geq 0$ we have $c_n > 0$, directly follows from the definition of the sequence c_n, but one can also see this immediately from the proof of part i) of Lemma 12.2.2. Finally, part iii) of the same lemma gives that $\gcd(b_n, c_n) = 1$, so the fraction r_n cannot be simplified. $\square$

Because of their importance, the fractions r_n hold a particular name.

Definition 12.2.3 *The fraction r_n $(n \geq 0)$ appearing in the previous theorem is called the n-th convergent of α.*

The next lemma opens up the way towards proving several important approximation properties of the convergents r_n.

Lemma 12.2.3 *Using the previous notation, for any $n \geq 0$*

$$|\alpha - r_n| < \frac{1}{c_n c_{n+1}}$$

holds.

Proof. By Lemmas 12.2.1 and 12.2.2 we obtain that for any $n \geq 0$

$$\alpha = [a_0; a_1, \ldots, a_n, \alpha_{n+1}] = \frac{\alpha_{n+1} b_n + b_{n-1}}{\alpha_{n+1} c_n + c_{n-1}}$$

holds. Hence

$$|\alpha - r_n| = \left|\alpha - \frac{b_n}{c_n}\right| = \left|\frac{\alpha_{n+1} b_n + b_{n-1}}{\alpha_{n+1} c_n + c_{n-1}} - \frac{b_n}{c_n}\right| = \left|\frac{b_{n-1} c_n - b_n c_{n-1}}{c_n(\alpha_{n+1} c_n + c_{n-1})}\right|.$$

Using part ii) of Theorem 12.2.1, we have

$$b_{n-1} c_n - b_n c_{n-1} = (-1)^n.$$

Thus by the above equalities, using $a_{n+1} = \lfloor \alpha_{n+1} \rfloor$ we get

$$|\alpha - r_n| = \frac{1}{c_n(\alpha_{n+1} c_n + c_{n-1})} < \frac{1}{c_n(a_{n+1} c_n + c_{n-1})} = \frac{1}{c_n c_{n+1}}.$$

This proves the statement. □

A simple but important consequence of the above lemma is that α is the limit of the sequence r_n.

Theorem 12.2.3 *By the above notation, we have*

$$\lim_{n \to \infty} r_n = \alpha.$$

Proof. By Lemma 12.2.3 we have

$$|\alpha - r_n| < \frac{1}{c_n c_{n+1}}.$$

Using part i) of Theorem 12.2.1, we obtain that

$$\lim_{n \to \infty} \frac{1}{c_n c_{n+1}} = 0.$$

From this our statement immediately follows. □

Remark 12.2.3 The above theorem shows a very important fact: the irrational number α is uniquely determined by the infinite continued fraction $[a_0; a_1, a_2, \dots]$. Formally, we can express this assertion as

$$[a_0; a_1, a_2, \dots] = \lim_{n\to\infty} [a_0; a_1, \dots, a_n],$$

or as

$$\alpha = [a_0; a_1, a_2, \dots].$$

Thus we may say that $[a_0; a_1, a_2, \dots]$ is the continued fraction expansion of α.

Theorem 12.2.4 *The fractions r_n approximate α in order two.*

Proof. The statement easily follows from Lemma 12.2.3, as for any $n \geq 0$ we have

$$c_{n+1} \geq c_n,$$

thus for any $n \geq 0$

$$|\alpha - r_n| < \frac{1}{c_n c_{n+1}} \leq \frac{1}{c_n^2}$$

holds. □

Remark 12.2.4 From the above statement we immediately conclude that the irrational numbers can be approximated in order two, and one can use the convergents as approximating fractions. In fact, the reverse statement is also 'almost' valid. Namely, one can prove that if α is an irrational number, and p/q ($p, q \in \mathbb{Z}$, $q > 0$, $\gcd(p, q) = 1$) is a rational number for which

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2},$$

then

$$\frac{p}{q} = r_n$$

holds with some $n \geq 0$.

12.3 Continued fraction factorization

In this section we give an important and interesting application of continued fractions for prime factorization. In fact, by their help we can improve upon the factorbase factorization algorithm. For this, we shall need the following two simple assertions.

Lemma 12.3.1 *Let α be an irrational number, $\alpha > 1$, and let b_n/c_n $(n \geq 0)$ be the convergents of α. Then for any $n \geq 0$ we have*

$$b_n^2 - \alpha^2 c_n^2 < 2\alpha.$$

Proof. The statement is a simple consequence of the following equalities and inequalities based upon the earlier discussed properties of the sequences b_n and c_n, and on Lemma 12.2.3:

$$\begin{aligned} |b_n^2 - \alpha^2 c_n^2| = |b_n - \alpha c_n| \cdot |b_n + \alpha c_n| = c_n^2 \cdot |\alpha - \frac{b_n}{c_n}| \cdot |\alpha + \frac{b_n}{c_n}| < \\ < \frac{c_n^2}{c_n c_{n+1}} \left| 2\alpha + \frac{b_n}{c_n} - \alpha \right| < \frac{c_n^2}{c_n c_{n+1}} \left(2\alpha + \frac{1}{c_n c_{n+1}} \right) = \\ = 2\alpha \left(\frac{c_n^2}{c_n c_{n+1}} + \frac{1}{2\alpha c_n c_{n+1}} \right) < 2\alpha \left(\frac{c_n}{c_{n+1}} + \frac{1}{c_{n+1}} \right) \leq 2\alpha. \end{aligned}$$

□

The next theorem is just a simple application of the above lemma. Before its formulation we recall our earlier notation, that if $m > 1$ is odd and $a \in \mathbb{Z}$, then $a \pmod m$ is the uniquely determined integer b, for which

$$a \equiv b \pmod m \quad \text{and} \quad -\frac{m}{2} < b < \frac{m}{2}$$

hold.

Theorem 12.3.1 *Let n be an odd, non-square positive integer and let b_i/c_i be the i-th convergent of the irrational number $\sqrt{n}$. Then for any $i \geq 0$ we have*

$$|b_i^2 \pmod n| < 2\sqrt{n}.$$

Proof. Using Lemma 12.3.1 with the choice $\alpha = \sqrt{n}$, we obtain

$$|b_i^2 - nc_i^2| < 2\sqrt{n} \quad (i \geq 0).$$

As

$$b_i^2 - nc_i^2 \equiv b_i^2 \pmod{n} \quad (i \geq 0),$$

this immediately implies our statement. □

Remark 12.3.1 By the above theorem we see that instead of integers b_i chosen randomly, it is better to take the numerators of the convergents of $\sqrt{n}$ in the factorbase factorization, where n is the integer to be factorized. Indeed, if b_i is chosen randomly then we can only guarantee

$$|b_i^2 \pmod{n}| < n/2,$$

while with the above choice we have

$$|b_i^2 \pmod{n}| < 2\sqrt{n}.$$

Thus we have much better chances for that the number b_i^2 is a B-integer. The bit operation requirement of the procedure after this modification becomes $O(e^{\sqrt{2 \log n \log \log n}})$.

We illustrate the above principle with the following example.

Example 12.3.1 Let the number to be factorized be given by

$$n = 9073,$$

and choose

$$B = \{-1, 2, 3, 7\}$$

as our factorbase. The continued fraction expansion of $\sqrt{n}$ is

$$\sqrt{9073} = [95; 3, 1, 26, 2, \dots],$$

whence (by the usual notation)

i	0	1	2	3	4
b_i	95	286	381	1119	2619
$b_i^2 \pmod{9073}$	-48	139	-7	87	-27

follows. Thus we see that for $i = 0, 2, 4$, b_i is a B-integer. We also get that

$$b_0^2 \pmod{9073} \cdot b_4^2 \pmod{9073}$$

is a square. From this we obtain

$$3834^2 \equiv 95^2 \cdot 2619^2 \equiv (-1)^2 \cdot 2^4 \cdot 3^4 \equiv 36^2 \pmod{9073}.$$

Since

$$3834^2 \not\equiv \pm 36 \pmod{9073},$$

thus based upon

$$\gcd(3834 + 36, 9073) = 43$$

and

$$\gcd(3834 - 36, 9073) = 211$$

we get the factorization

$$9073 = 43 \cdot 211.$$

12.4 Exercises

Exercise 12.4.1 Give the continued fraction representations of the following rational numbers:

- $\frac{53}{11}$
- $\frac{-29}{13}$
- $\frac{101}{29}$
- $\frac{-101}{203}$
- $\frac{77}{89}$

Exercise 12.4.2 Determine the rational numbers with the following continued fraction representations:

- $[1; 2, 3, 4]$

- $[-1; 1, 1, 2, 2]$
- $[0; 1, 3, 5, 7]$
- $[1; 0, 1, 4, 7, 11]$

Exercise 12.4.3 Find the continued fraction expansions of the following irrational numbers:

- $\sqrt{2}$
- $\frac{1+\sqrt{5}}{2}$
- $\sqrt{3}$
- $-\sqrt{2}$
- $\sqrt{m^2+1}$ $(m \geq 1)$
- $\sqrt{m^2-1}$ $(m \geq 2)$

Exercise 12.4.4 Find the irrational numbers whose continued fraction expansions are the following:

- $[1; 2, 1, 2, 1, \dots]$
- $[2; 1, 2, 1, 2, \dots]$
- $[0; 3, 3, 5, 5, 3, 3, 5, 5, \dots]$
- $[-1; 1, 2, 3, 1, 2, 3, \dots]$

Exercise 12.4.5 Give the first four convergents of the following irrational numbers:

- $\sqrt{2}$
- $\frac{1+\sqrt{5}}{2}$
- π
- e

Exercise 12.4.6 Prove that for any $n \geq 3$ we have

$$\frac{F_n}{F_{n-1}} = [1, \ldots, 1, 2] = [1, \ldots, 1, 1, 1],$$

where in the first continued fractions the number of digits 1 is $n-3$ (and in the second one the number of digits 1 is $n-1$). Here F_n is the n-th Fibonacci-number, that is $F_0 = 0$, $F_1 = 1$ and

$$F_n = F_{n-1} + F_{n-2} \quad (n \geq 2).$$

Exercise 12.4.7 Let α be an irrational number, and let $[a_0; a_1, a_2, \ldots]$ be the continued fraction expansion of α. Prove that this continued fraction is periodic if and only if α is a quadratic algebraic number! (The above infinite continued fraction is called periodic if there exists a positive integer k and a non-negative integer n_0, such that $a_{n+k} = a_n$ holds for every $n \geq n_0$.)

Exercise 12.4.8 Factorize the number $n = 17873$ by the help of continued fraction factorization! For this, use the factorbase $B = \{-1, 2, 7, 23\}$ and the convergents of $\sqrt{n}$ of indices $i = 0, 2, 4, 5$.

Chapter 13

The LLL algorithm and its applications

In this chapter we study a very efficient algorithm, having several applications of many different types, namely, the so-called Lenstra-Lenstra-Lovász (shortly LLL) algorithm, named after its inventors. We also discuss some applications.

13.1 Lattices and LLL reduced bases

In this section we give the background of the LLL algorithm, together with some of its properties.

First we recall the Gram-Schmidt orthogonalization procedure, because this (in particular, orthogonalized bases and their properties) play an important role later on. As the result is well-known, we do not give its proof.

Theorem 13.1.1 (Gram-Schmidt orthogonalization) *Let $b_1, \ldots, b_n$ be a basis of $\mathbb{R}^n$. Then there exists a basis $b_1^*, \ldots, b_n^*$ of $\mathbb{R}^n$, for which*

$$b_i^* = b_i - \sum_{j=1}^{i-1} \mu_{i,j} b_j^* \quad (i = 1, \ldots, n)$$

holds. Here

$$\mu_{i,j} = \frac{(b_i, b_j^*)}{(b_j^*, b_j^*)} \quad (1 \leq j < i \leq n),$$

where $(.,.)$ is the standard inner product in $\mathbb{R}^n$.

The next notion is of utmost importance.

Definition 13.1.1 *Let Λ be a full lattice in $\mathbb{R}^n$, and $b_1, \ldots, b_n$ a basis of Λ. If (with the notation of Theorem 13.1.1) the following properties hold:*

1) $|\mu_{i,j}| \leq \frac{1}{2}$ $(1 \leq j < i \leq n)$,

2) $|b_i^* + \mu_{i,i-1} b_i^*|^2 \geq \frac{3}{4} |b_{i-1}^*|^2$ $(1 < i \leq n)$,

then we say that the basis $b_1, \ldots, b_n$ is LLL-reduced.

Remark 13.1.1 At this point we mention three important things.

1) The above point 1) shows that the vectors in an LLL-reduced basis are 'almost orthogonal'.

2) The above point 2) gives a condition for the relative lengths of the basis vectors. Indeed, this inequality, in view of that the vectors b_i^* $(i = 1, \ldots, n)$ are orthogonal, can be written in the shape

$$|b_i^*|^2 \geq \left(\frac{3}{4} - \mu_{i,i-1}^2 \right) |b_{i-1}^*|^2.$$

3) Every lattice Λ possesses an LLL-reduced basis, and such a basis can be obtained by a polynomial algorithm. We do not outline the algorithm here.

The following theorem gives several properties of the LLL-reduced bases.

Theorem 13.1.2 *Let Λ be a full lattice in $\mathbb{R}^n$, $b_1, \ldots, b_n$ an LLL-reduced basis of Λ, and $b_1^*, \ldots, b_n^*$ the Gram-Schmidt basis obtained from this by Theorem 13.1.1. Then the following assertions are valid:*

i) $|b_j|^2 \leq 2^{i-1} |b_i|^2$ $(1 \leq j \leq i \leq n)$,

ii) $\det(\Lambda) \leq \prod_{i=1}^{n} |b_i| \leq 2^{n(n-1)/4} \det(\Lambda)$,

iii) $|b_1| \leq 2^{(n-1)/4} (\det(\Lambda))^{1/n}$.

Proof. We prove the three statements separately.

i) Using that the basis $b_1, \ldots, b_n$ is LLL-reduced, for $i = 2, \ldots, n$ we obtain that

$$|b_i^*|^2 \geq \left(\frac{3}{4} - \mu_{i,i-1}^2\right) |b_{i-1}^*|^2 \geq \frac{1}{2}|b_i^*|^2.$$

(See point 2) of the previous Remark.) From this for $1 \leq j \leq i \leq n$ by induction on j the inequality

$$|b_j^*|^2 \leq 2^{i-j}|b_i^*|^2$$

easily follows. Hence by using the properties of the Gram-Schmidt basis we get

$$|b_i|^2 = |b_i^*|^2 + \sum_{j=1}^{i-1} \mu_{i,j}^2 |b_j^*|^2 \leq \left(1 + \sum_{j=1}^{i-1} 2^{i-j-2}\right) |b_i^*|^2 =$$
$$= \left(1 + \frac{1}{4}(2^i - 2)\right) |b_i^*|^2 \leq 2^{i-1}|b_i^*|^2.$$

This proves our claim.

ii) Observe that the basis transformation matrix between the vector systems $b_1, \ldots, b_n$ and $b_1^*, \ldots, b_n^*$, as bases of $\mathbb{R}^n$, is a triangular matrix having only 1-s in its main diagonal. Thus

$$\det(b_1, \ldots, b_n) = \det(b_1^*, \ldots, b_n^*),$$

where the above determinants belong to matrices obtained from the basis vectors as column vectors in $\mathbb{R}^n$. As $b_1^*, \ldots, b_n^*$ is an orthogonal vector system, hence

$$\det(\Lambda) = |\det(b_1, \ldots, b_n)| = |\det(b_1^*, \ldots, b_n^*)| = \prod_{i=1}^{n} |b_i^*|$$

follows. Using the already verified part i) of the theorem and the trivial assertion

$$|b_i^*| \leq |b_i| \quad (i = 1, \ldots, n),$$

we obtain that

$$\det(\Lambda) \leq \prod_{i=1}^{n} |b_i| \leq \prod_{i=1}^{n} 2^{(i-1)/2}|b_i^*| \leq$$
$$\leq 2^{n(n-1)/4} \prod_{i=1}^{n} |b_i^*| = 2^{n(n-1)/4} \det(\Lambda).$$

This proves our claim.

iii) Applying the already proved statement i) with $j = 1$, we get

$$|b_1|^2 = 2^{i-1}|b_i^*|^2 \quad (i = 1, \ldots, n).$$

Taking the products of the above left- and right hand sides for $i = 1, \ldots, n$, using the assertion

$$\prod_{i=1}^{n} |b_i^*| = \det(\Lambda),$$

we get that

$$|b_i|^{2n} \leq \prod_{i=1}^{n} 2^{i-1}|b_i^*|^2 \leq 2^{n(n-1)/2} \det(\Lambda)^2.$$

From this our statement directly follows. □

Our next theorem shows that the length of the first vector of an LLL-reduced basis is not 'much larger' then the length of the shortest (non-zero) vector of the lattice.

Theorem 13.1.3 *Let $b_1, \ldots, b_n$ be an LLL-reduced basis of a full lattice Λ. Then for any non-zero vector x of Λ we have*

$$|b_1|^2 \leq c_1|x|^2,$$

where

$$c_1 = \max_{1 \leq i \leq n} \frac{|b_1|^2}{|b_i^*|^2}.$$

Proof. First, observe that

$$|b_1|^2 \leq c_1|b_i^*|^2 \quad (i = 1, \ldots, n).$$

Let

$$x = \sum_{i=1}^{n} r_i b_i = \sum_{i=1}^{n} r_i' b_i^*,$$

where $r_i \in \mathbb{Z}$, $r_i' \in \mathbb{R}$ $(i = 1, \ldots, n)$. Let i_0 be the largest index i, for which $r_i \neq 0$ holds. (Since $x \neq 0$, such an index exists.) Then by Theorem 13.1.1 we see that

$$r_{i_0}' = r_{i_0},$$

whence

$$|x|^2 = \sum_{i=1}^{n} r_i'^2 |b_i^*|^2 \geq r_{i_0}'^2 |b_{i_0}^*|^2 \geq |b_{i_0}^*|^2 \geq \frac{1}{c_1}|b_1|^2.$$

Hence our statement follows. □

Remark 13.1.2 By part i) of Theorem 13.1.2 we get that the above theorem also holds with the choice $c_1 = 2^{n-1}$.

13.2 Approximation lattices

In this section we build a kind of bridge between the LLL algorithm and its applications: we introduce and study approximation lattices.

Our starting point (which is the starting point of many applications, as well) is finding approximate solutions of a linear form. First introduce the following notion.

Definition 13.2.1 *Let $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Then*

$$L(x_1, \ldots, x_n) = \alpha_0 + x_1\alpha_1 + \cdots + x_n\alpha_n$$

is a linear form. If $\alpha_0 = 0$, then we say that the linear form is homogeneous; otherwise it is inhomogeneous.

Remark 13.2.1 As we have already mentioned, our basic question is that how 'small' can be the value of

$$|L(x_1, \ldots, x_n)|,$$

such that

$$\max_{i=1,\ldots,n} |x_i|$$

is not 'too large'. For the sake of simplicity we may clearly assume that

- in the homogeneous case not all the x_i $(i = 1, \ldots, n)$ are zero,
- the numbers $\alpha_0, \alpha_1, \ldots, \alpha_n$ are linearly independent over $\mathbb{Z}$, that is, in case of

  $$L(x_1, \ldots, x_n) = 0$$

  we have

  $$x_1 = \cdots = x_n = 0.$$

In the following we shall use these assumptions without any further mentioning.

Now we introduce the notion of the approximation lattice. It will be of great help to achieve our aims formulated in the previous Remark.

Definition 13.2.2 *With the previous notation, consider the $n \times n$ type real matrix*

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \alpha_1 & \alpha_2 & \dots & \alpha_{n-1} & \alpha_n \end{pmatrix}.$$

The full lattice Λ generated by the column vectors of A in $\mathbb{R}^n$ is called an approximation lattice.

Remark 13.2.2 Consider the vectors of the form Ax, where $x \in \mathbb{Z}^n$. Our purpose is to find a 'short' vector x_0, for which Ax_0 is 'close' to the vector

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\alpha_0 \end{pmatrix}$$

in $\mathbb{R}^n$. In this way we succeeded to translate or Diophantine approximation problem, that is, the minimization problem of $|L(x_1, \dots, x_n)|$, into a problem concerning lattices. Thus we may use the LLL algorithm and LLL-reduced bases. We shall present applications in the following section.

Observe that if the values α_i $(i = 1, \dots, n)$ are 'small', then the lattice Λ consists of vectors with 'small' n-th component with respect to any of the first $n-1$ ones. Hence it seems to be worth to re-scale the approximation lattice (which works quite well in practice). For this, choose a 'large' integer C, and in place of Λ consider the lattice Λ_C generated by the column vectors of the matrix

$$A_C = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ ||C\alpha_1|| & ||C\alpha_2|| & \dots & ||C\alpha_{n-1}|| & ||C\alpha_n|| \end{pmatrix}.$$

Here $||C\alpha_i||$ denotes the integer closest to $C\alpha_i$; if $2C\alpha_i \in \mathbb{Z}$ then $||C\alpha_i|| = \lfloor C\alpha_i \rfloor$ $(i = 1, \ldots, n)$. Obviously, the principle outlined above is valid for the lattice Λ_C, as well.

Finally, we mention that our method can be extended to the case where instead of one linear form we have a system

$$L_j(x_1, \ldots, x_n) = \alpha_{j,0} + x_1\alpha_{j,1} + \cdots + x_n\alpha_{j,n} \quad (1 \leq j \leq m)$$

consisting of linear forms, where $m \leq n$. In this case we may use the approximation lattices generated by the column vectors of the following $n \times n$ type real matrix:

$$\begin{pmatrix} 1 & & & 0 & \\ & \ddots & & & \\ 0 & & 1 & & \\ ||C\alpha_{1,1}|| & \ldots & ||C\alpha_{1,n-m}|| & \ldots & ||C\alpha_{1,n}|| \\ \vdots & & \vdots & & \vdots \\ ||C\alpha_{m,1}|| & \ldots & ||C\alpha_{m,n-m}|| & \ldots & ||C\alpha_{m,n}|| \end{pmatrix}.$$

13.3 Applications

In this section we demonstrate the applicability of the LLL algorithm through two examples.

Example 13.3.1 Find a polynomial with integer coefficients of degree at most three, such that for some of its roots α

$$|\alpha - \pi| < 0.01$$

holds!

To solve the problem, we need to find integers x_1, x_2, x_3, x_4 such that the value of the linear form

$$L(x_1, x_2, x_3, x_4) = x_1\pi^3 + x_2\pi^2 + x_3\pi + x_4$$

is 'small', more precisely,

$$|L(x_1, x_2, x_3, x_4)| < 0.01$$

holds. For this it seems to be appropriate to find values x_i $(i = 1, 2, 3, 4)$ of magnitude 10^2. (As we shall see, this expectation will prove to be valid.) So consider the approximation lattice Λ generated by the column vectors of the following matrix:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 3101 & 987 & 314 & 100 \end{pmatrix},$$

where the entries of the last row are the numbers

$$||100\pi^3||,\ ||100\pi^2||,\ ||100\pi^1||,\ ||100\pi^0||,$$

respectively. By our previous discussion it is clear that the shortest non-zero vector of this lattice Λ provides a polynomial of degree (at most) three, whose coefficients are 'not too large', and one of its roots is 'close' to π. By the help of a computer (e.g. using the program package Maple) find an LLL-reduced basis of the lattice Λ. Such a basis is given by the column vectors of the matrix AU, where

$$U = \begin{pmatrix} -1 & -1 & -2 & -3 \\ 0 & 1 & 2 & -6 \\ 0 & 1 & -5 & 2 \\ 31 & 18 & 58 & 146 \end{pmatrix}.$$

From the first column of U we obtain that one of the roots α of the polynomial $x^3 - 31$ is 'close' to π. Indeed, here

$$|\alpha - \pi| \leq 0.00022$$

holds.

Modify the problem such that we are looking for a polynomial of degree at most three such that for two roots α and β of the polynomial

$$|\alpha - \pi| \leq 0.01 \quad \text{and} \quad |\beta - e| < 0.01$$

holds!

In this case the approximation lattice Λ is chosen as the lattice generated by the column vectors of the following matrix:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 3101 & 987 & 314 & 100 \\ 2009 & 739 & 272 & 100 \end{pmatrix}.$$

Here the entries of the third row are given by

$$||100\pi^3||,\ ||100\pi^2||,\ ||100\pi^1||,\ ||100\pi^0||,$$

respectively, while the entries of the fourth row are

$$||100e^3||,\ ||100e^2||,\ ||100e^1||,\ ||100e^0||.$$

Using the above notation, we get

$$U = \begin{pmatrix} 3 & 0 & 8 & 1 \\ -2 & 1 & 1 & -9 \\ -66 & -6 & -214 & 27 \\ 134 & 9 & 414 & -27 \end{pmatrix}.$$

From the first column of U we obtain that two roots α, β of the polynomial $3x^3 - 2x^2 - 66x + 134$ are 'close' to π and e, respectively. Indeed, here

$$|\alpha - \pi| \leq 0.007 \quad \text{and} \quad |\beta - e| \leq 0.008$$

hold.

Example 13.3.2 Find a linear relation among the numbers

$$\arctan(1), \quad \arctan\left(\frac{1}{5}\right), \quad \arctan\left(\frac{1}{239}\right).$$

To solve the problem, find integers x_1, x_2, x_3 such that the value

$$L(x_1, x_2, x_3) = x_1 \arctan(1) + x_2 \arctan\left(\frac{1}{5}\right) + x_3 \arctan\left(\frac{1}{239}\right)$$

is 'small'. For this, let Λ be the approximation lattice generated by the column vectors of the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 78540 & 19740 & 418 \end{pmatrix}.$$

The entries of the last row are the following:

$$||10^5 \arctan(1)||, \quad \left|\left|10^5 \arctan\left(\frac{1}{5}\right)\right|\right|, \quad \left|\left|10^5 \arctan\left(\frac{1}{239}\right)\right|\right|.$$

As an LLL-reduced basis of Λ we get AU, where

$$U = \begin{pmatrix} 1 & 3 & 13 \\ -4 & -3 & 6 \\ 1 & -422 & -2726 \end{pmatrix}.$$

From this, we 'conjecture' the values $x_1 = 1$, $x_2 = -4$, $x_3 = 1$. This is supported by the assertion

$$\arctan(1) - 4\arctan\left(\frac{1}{5}\right) + \arctan\left(\frac{1}{239}\right) \approx 0.0000000000302$$

obtained by 10 digit precision. Having the 'conjecture', it is easy to check that

$$\arctan(1) - 4\arctan\left(\frac{1}{5}\right) + \arctan\left(\frac{1}{239}\right) = 0 \tag{13.1}$$

is an identity. Indeed, using the well-known formula

$$e^{2i\arctan t} = \frac{1+it}{1-it} \quad \left(-\frac{\pi}{2} < t < \frac{\pi}{2}\right),$$

the assertion (13.1) is equivalent to the following:

$$\frac{(1+i)(1+5i)^{-4}(1+239i)}{(1-i)(1-5i)^{-4}(1-239i)} = 1,$$

which is easy to check. We mention that (13.1) is the so-called Machin formula.

13.4 Exercises

Exercise 13.4.1 Find a polynomial with integer coefficients of degree at most three, such that for some of its roots α

$$|\alpha - \sqrt[4]{2}| < 0.0001$$

holds!

Exercise 13.4.2 Find a polynomial with integer coefficients of degree at most three, such that for some of its roots α, β

$$|\alpha - \sqrt[4]{2}| < 0.01 \quad \text{and} \quad |\beta - \sqrt[4]{3}| < 0.01$$

holds!

Exercise 13.4.3 Find all integer solutions x, y, z of the inequality

$$|x \log 2 + y \log 3 - z \log 5| < 0.01$$

for which

$$\max(|x|, |y|, |z|) < 10^5$$

holds!

References

1. T. M. Apostol, *Introduction to Analytic Number Theory*, Springer, 1976.

2. P. Erdős, J. Surányi, *Topics in the Theory of Numbers*, Undergraduate Texts in Mathematics, Springer, 2003.

3. Freud R., Gyarmati E., *Számelmélet*, Nemzeti Tankönyvkiadó, Budapest, 2000.

4. N. Koblitz, *A Course in Number Theory and Cryptography*, Springer, 1994.

5. Maple 16. *Maplesoft, a division of Waterloo Maple Inc.*, Waterloo, Ontario.

6. I. Niven, H. S. Zuckerman, H. L. Montgomery, *An introduction to the theory of numbers*, 5th edition, Wiley, 1991.

7. Sárközy A., Surányi J., *Számelmélet feladatgyűjtemény*, Tankönyvkiadó, Budapest, 1986.

8. N. P. Smart, *The Algorithmic Resolution of Diophantine Equations*, London Mathematical Society Student Texts 41, Cambridge University Press, Cambridge, 1998.