1. Algebra

First, in this chapter, we will discuss some of the basic concepts of algebra, such as fields, vector spaces and polynomials (Section 1.1). Our main focus will be the study of polynomial rings in one variable. These polynomial rings play a very important rôle in *constructive applications*. After this, we will outline the theory of finite fields, putting a strong emphasis on the problem of constructing them (Section 1.2) and on the problem of factoring polynomials over such fields (Section 1.3). Then we will study lattices and discuss the Lenstra-Lenstra-Lovász algorithm which can be used to find short lattice vectors (Section 1.4). We will present a polynomial time algorithm for the factorisation of polynomials with rational coefficients; this was the first notable application of the Lenstra-Lenstra-Lovász algorithm (Section 1.5).

1.1. Fields, vector spaces, and polynomials

In this section we will overview some important concepts related to rings and polynomials.

1.1.1. Ring theoretic concepts

We recall some definitions introduced in Chapters 31-33 of the textbook *Introduction to Algorithms*. In the sequel all cross references to Chapters 31-33 refer to results in that book.

Reference to New Algorithms.

A set S with at least two elements is called a ***ring***, if it has two binary operations, the addition, denoted by the + sign, and the multiplication, denoted by the · sign. The elements of S form an abelian group with respect to the addition, and they form a monoid (that is, a semigroup with an identity), whose identity element is denoted by 1, with respect to the multiplication. We assume that $1 \neq 0$. Further, the distributive properties also hold: for arbitrary elements $a, b, c \in S$ we have

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

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

Being an abelian group with respect to the addition means that the operation is associative, commutative, it has an identity element (denoted by 0), and every element has an inverse with respect to this identity. More precisely, these requirements are the following:
associative property: for all triples $a,\ b,\ c \in S$ we have $(a + b) + c = a + (b + c)$;
commutative property: for all pairs $a,\ b \in S$ we have $a + b = b + a$;
existence of the identity element: for the zero element 0 of S and for all elements a of S, we have $a + 0 = 0 + a = a$;
existence of the additive inverse: for all $a \in S$ there exists $b \in S$, such that $a + b = 0$.
It is easy to show that each of the elements a in S has a unique inverse. We usually denote the inverse of an element a by $-a$.

Concerning the multiplication, we require that it must be associative and that the multiplicative identity should exist. The ***identity*** of a ring S is the multiplicative identity of S. The usual name of the additive identity is ***zero***. We usually omit the $\cdot$ sign when writing the multiplication, for example we usually write ab instead of $a \cdot b$.

1.1. Example. *Rings.*
(i) The set $\mathbf{Z}$ of integers with the usual operations + and $\cdot$.
(ii) The set $\mathbf{Z}_m$ of residue classes modulo m with respect to the addition and multiplication modulo m.
(iii) The set $\mathbb{R}^{n\times n}$ of $(n \times n)$-matrices with real entries with respect to the addition and multiplication of matrices.

Let S_1 and S_2 be rings. A map $\phi : S_1 \to S_2$ is said to be a ***homomorphism***, if ϕ preserves the operations, in the sense that $\phi(a \pm b) = \phi(a) \pm \phi(b)$ and $\phi(ab) = \phi(a)\phi(b)$ holds for all pairs $a,\ b \in S_1$. A homomorphism ϕ is called an ***isomorphism***, if ϕ is a one-to-one correspondence, and the inverse is also a homomorphism. We say that the rings S_1 and S_2 are ***isomorphic***, if there is an isomorphism between them. If S_1 and S_2 are isomorphic rings, then we write $S_1 \cong S_2$. From an algebraic point of view, isomorphic rings can be viewed as identical.

For example the map $\phi : \mathbf{Z} \to \mathbf{Z}_6$ which maps an integer to its residue modulo 6 is a homomorphism: $\phi(13) = 1,\ \phi(5) = 5,\ \phi(22) = 4$, etc.

A useful and important ring theoretic construction is the ***direct sum.*** The direct sum of the rings S_1 and S_2 is denoted by $S_1 \oplus S_2$. The underlying set of the direct sum is $S_1 \times S_2$, that is, the set of ordered pairs (s_1, s_2) where $s_i \in S_i$. The operations are defined componentwise: for $s_i,\ t_i \in S_i$ we let

$$(s_1, s_2) + (t_1, t_2) := (s_1 + t_1, s_2 + t_2) \quad \text{and}$$

$$(s_1, s_2) \cdot (t_1, t_2) := (s_1 \cdot t_1, s_2 \cdot t_2)\ .$$

Easy calculation shows that $S_1 \oplus S_2$ is a ring with respect to the operations above. This construction can easily be generalised to more than two rings. In this case, the elements of the direct sum are the k-tuples, where k is the number of rings in the direct sum, and the operations are defined componentwise.

Fields

A ring $\mathbb{F}$ is said to be a ***field***, if its non-zero elements form an abelian group with respect to the multiplication. The multiplicative inverse of a non-zero element a is usually denoted a^{-1}.

The best-known examples of fields are the the sets of rational numbers, real numbers,

and complex numbers with respect to the usual operations. We usually denote these fields by $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, respectively.

Another important class of fields consists of the fields $\mathbb{F}_p$ of p-elements where p is a prime number. The elements of $\mathbb{F}_p$ are the residue classes modulo p, and the operations are the addition and the multiplication defined on the residue classes. The distributive property can easily be derived from the distributivity of the integer operations. By Theorem 33.12, $\mathbb{F}_p$ is a group with respect to the addition, and, by Theorem 33.13, the set $\mathbb{F}_p^*$ of non-zero elements of $\mathbb{F}_p$ is a group with respect to the multiplication. In order to prove this latter claim, we need to use that p is a prime number.

Reference to NA!

Characteristic, prime field

In an arbitrary field, we may consider the set of elements of the form $m \cdot 1$, that is, the set of elements that can be written as the sum $1 + \cdots + 1$ of m copies of the multiplicative identity where m is a positive integer. Clearly, one of the two possibilities must hold:
(a) none of the elements $m \cdot 1$ is zero;
(b) $m \cdot 1$ is zero for some $m \geq 1$.

In case (a) we say that $\mathbb{F}$ is a field with ***characteristic zero***. In case (b) the ***characteristic*** of $\mathbb{F}$ is the smallest $m \geq 1$ such that $m \cdot 1 = 0$. In this case, the number m must be a prime, for, if $m = rs$, then $0 = m \cdot 1 = rs \cdot 1 = (r \cdot 1)(s \cdot 1)$, and so either $r \cdot 1 = 0$ or $s \cdot 1 = 0$.

Suppose that P denotes the smallest subfield of $\mathbb{F}$ that contains 1. Then P is said to be the ***prime field*** of $\mathbb{F}$. In case (a) the subfield P consists of the elements $(m \cdot 1)(s \cdot 1)^{-1}$ where m is an integer and s is a positive integer. In this case, P is isomorphic to the field $\mathbb{Q}$ of rational numbers. The identification is obvious: $(m \cdot 1)(s \cdot 1)^{-1} \leftrightarrow m/s$.

In case (b) the characteristic is a prime number, and P is the set of elements $m \cdot 1$ where $0 \leq m < p$. In this case, P is isomorphic to the field $\mathbb{F}_p$ of residue classes modulo p.

Vector spaces

Let $\mathbb{F}$ be a field. An additively written abelian group V is said to be a ***vector space*** over $\mathbb{F}$, or simply an $\mathbb{F}$-vector space, if for all elements $a \in \mathbb{F}$ and $v \in V$, an element $av \in V$ is defined (in other words, $\mathbb{F}$ acts on V) and the following hold:

$$a(u + v) = au + av, \quad (a + b)u = au + bu \,,$$

$$a(bu) = (ab)u, \quad 1u = u \,.$$

Here a, b are arbitrary elements of $\mathbb{F}$, the elements u, v are arbitrary in V, and the element 1 is the multiplicative identity of $\mathbb{F}$.

The space of $(m \times n)$-matrices over $\mathbb{F}$ is an important example of vector spaces. Their properties are studied in Chapter 31.

reference to NA

A vector space V over a field $\mathbb{F}$ is said to be ***finite-dimensional*** if there is a collection $\{v_1, \ldots, v_n\}$ of finitely many elements in V such that each of the elements $v \in V$ can be written as a ***linear combination*** $v = a_1 v_1 + \cdots + a_n v_n$ for some $a_1, \ldots, a_n \in \mathbb{F}$. Such a set $\{v_i\}$ is called a ***generating set*** of V. The cardinality of the smallest generating set of V is referred to as the ***dimension*** of V over $\mathbb{F}$, denoted $\dim_{\mathbb{F}} V$. In a finite-dimensional vector space, a generating system containing $\dim_{\mathbb{F}} V$ elements is said to be a ***basis***.

A set $\{v_1, \ldots, v_k\}$ of elements of a vector space V is said to be ***linearly independent***, if, for $a_1, \ldots, a_k \in \mathbb{F}$, the equation $0 = a_1 v_1 + \cdots + a_k v_k$ implies $a_1 = \cdots = a_k = 0$.

It is easy to show that a basis in V is a linearly independent set. An important property of linearly independent sets is that such a set can be extended to a basis of the vector space. The dimension of a vector space coincides with the cardinality of its largest linearly independent set.

A non-empty subset U of a vector space V is said to be a ***subspace*** of V, if it is an (additive) subgroup of V, and $au \in U$ holds for all $a \in \mathbb{F}$ and $u \in U$. It is obvious that a subspace can be viewed as a vector space.

The concept of homomorphisms can be defined for vector spaces, but in this context we usually refer to them as ***linear maps***. Let V_1 and V_2 be vector spaces over a common field $\mathbb{F}$. A map $\phi : V_1 \to V_2$ is said to be linear, if, for all $a,\ b \in \mathbb{F}$ and $u,\ v \in V_1$, we have

$$\phi(au + bv) = a\phi(u) + b\phi(v) \ .$$

The linear mapping ϕ is an ***isomorphism*** if ϕ is a one-to-one correspondence and its inverse is also a homomorphism. Two vector spaces are said to be isomorphic if there is an isomorphism between them.

Lemma 1.1 *Suppose that $\phi : V_1 \to V_2$ is a linear mapping. Then $U = \phi(V_1)$ is a subspace in V_2. If ϕ is one-to-one, then $\dim_{\mathbb{F}} U = \dim_{\mathbb{F}} V_1$. If, in this case, $\dim_{\mathbb{F}} V_1 = \dim_{\mathbb{F}} V_2 < \infty$, then $U = V_2$ and the mapping ϕ is an isomorphism.*

Proof. As

$$\phi(u) \pm \phi(v) = \phi(u \pm v) \quad \text{and} \quad a\phi(u) = \phi(au),$$

we obtain that U is a subspace. Further, it is clear that the images of the elements of a generating set of V_1 form a generating set for U. Let us now suppose that ϕ is one-to-one. In this case, the image of a linearly independent subset of V_1 is linearly independent in V_2. It easily follows from these observations that the image of a basis of V_1 is a basis of U, and so $\dim_{\mathbb{F}} U = \dim_{\mathbb{F}} V_1$. If we assume, in addition, that $\dim_{\mathbb{F}} V_2 = \dim_{\mathbb{F}} V_1$, then a basis of U is also a basis of V_2, as it is a linearly independent set, and so it can be extended to a basis of V_2. Thus $U = V_2$ and the mapping ϕ must be a one-to-one correspondence. It is easy to see, and is left to the reader, that ϕ^{-1} is a linear mapping. ■

The ***direct sum*** of vector spaces can be defined similarly to the direct sum of rings. The direct sum of the vector spaces V_1 and V_2 is denoted by $V_1 \oplus V_2$. The underlying set of the direct sum is $V_1 \times V_2$, and the addition and the action of the field $\mathbb{F}$ are defined componentwise. It is easy to see that

$$\dim_{\mathbb{F}} (V_1 \oplus V_2) = \dim_{\mathbb{F}} V_1 + \dim_{\mathbb{F}} V_2 \ .$$

Finite multiplicative subgroups of fields

Let $\mathbb{F}$ be a field and let $G \subseteq \mathbb{F}$ be a finite multiplicative subgroup of $\mathbb{F}$. That is, the set G contains finitely many elements of $\mathbb{F}$, each of which is non-zero, G is closed under multiplication, and the multiplicative inverse of an element of G also lies in G. We aim to show that the group G is cyclic, that is, G can be generated by a single element. The main concepts related to cyclic groups can be found in Section 33.3.4. Recall that the order $\mathrm{ord}(a)$ of an element $a \in G$ is the smallest positive integer k such that $a^k = 1$.

Reference to NA

The cyclic group generated by an element a is denoted by $\langle a \rangle$. Clearly, $|\langle a \rangle| = \mathrm{ord}(a)$, and an element a^i generates the group $\langle a \rangle$ if and only if i and n are relatively prime. Hence

Reference to NA!

the group $\langle a \rangle$ has exactly $\phi(n)$ generators where ϕ is Euler's totient function (see 33.3.2).

The following identity is valid for an arbitrary integer n:

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

Here the summation index d runs through all positive divisors of n. In order to verify this identity, consider all the rational numbers i/n with $1 \leq i \leq n$. The number of these is exactly n. After simplifying these fractions, they will be of the form j/d where d is a positive divisor of n. A fixed denominator d will occur exactly $\phi(d)$ times.

Theorem 1.2 *Suppose that $\mathbb{F}$ is a field and let G be a finite multiplicative subgroup of $\mathbb{F}$. Then there exists an element $a \in G$ such that $G = \langle a \rangle$.*

Proof. Suppose that $|G| = n$. Lagrange's theorem (Theorem 33.15) implies that the order of an element $b \in G$ is a divisor of n. We claim, for an arbitrary d, that there are at most $\phi(d)$ elements in $\mathbb{F}$ with order d. The elements with order d are roots of the polynomial $x^d - 1$. If $\mathbb{F}$ has an element b with order d, then, by Lemma 1.5, $x^d - 1 = (x - b)(x - b^2) \cdots (x - b^d)$ (the lemma will be verified later). Therefore all the elements of $\mathbb{F}$ with order d are contained in the group $\langle b \rangle$, which, in turn, contains exactly $\phi(d)$ elements of order d.

Reference to NA!

If G had no element of order n, then the order of each of the elements of G would be a proper divisor of n. In this case, however, using the identity above and the fact that $\phi(n) > 0$, we obtain

$$n = |G| \leq \sum_{d|n,\ d<n} \phi(d) < n\ ,$$

which is a contradiction. ∎

1.1.2. Polynomials

Suppose that $\mathbb{F}$ is a field and that $a_0, \ldots, a_n$ are elements of $\mathbb{F}$. Recall that an expression of the form

$$f = f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where x is an indeterminate, is said to be a ***polynomial*** over $\mathbb{F}$ (see Chapter 32). The scalars a_i are the ***coefficients*** of the polynomial f. The degree of the zero polynomial is zero, while the ***degree*** of a non-zero polynomial f is the largest index j such that $a_j \neq 0$. The degree of f is denoted by $\deg f$.

Reference to NA!

The set of all polynomials over $\mathbb{F}$ in the indeterminate x is denoted by $\mathbb{F}[x]$. If

$$f = f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

and

$$g = g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n$$

are polynomials with degree not larger than n, then their sum is defined as the polynomial

$$h = h(x) = f + g = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n$$

whose coefficients are $c_i = a_i + b_i$.

The product fg of the polynomials f and g is defined as the polynomial

$$fg = d_0 + d_1x + d_2x^2 + \cdots + d_{2n}x^{2n}$$

with degree at most $2n$ whose coefficients are given by the equations $d_j = \sum_{k=0}^{j} a_k b_{j-k}$. On the right-hand side of these equations, the coefficients with index greater than n are considered zero. Easy computation shows that $\mathbb{F}[x]$ is a commutative ring with respect to these operations. It is also straightforward to show that $\mathbb{F}[x]$ has no ***zero divisors***, that is, whenever $fg = 0$, then either $f = 0$ or $g = 0$.

Division with remainder and divisibility

The ring $\mathbb{F}[x]$ of polynomials over $\mathbb{F}$ is quite similar, in many ways, to the ring $\mathbf{Z}$ of integers. One of their similar features is that the procedure of division with remainder can be performed in both rings.

Lemma 1.3 *Let $f(x),\ g(x) \in \mathbb{F}[x]$ be polynomials such that $g(x) \neq 0$. Then there there exist polynomials $q(x)$ and $r(x)$ such that*

$$f(x) = q(x)g(x) + r(x)\ ,$$

and either $r(x) = 0$ or $\deg\ r(x) < \deg\ g(x)$. Moreover, the polynomials q and r are uniquely determined by these conditions.

Proof. We verify the claim about the existence of the polynomials q and r by induction on the degree of f. If $f = 0$ or $\deg\ f < \deg\ g$, then the assertion clearly holds. Let us suppose, therefore, that $\deg\ f \geq \deg\ g$. Then subtracting a suitable multiple $q^*(x)g(x)$ of g from f, we obtain that the degree of $f_1(x) = f(x) - q^*(x)g(x)$ is smaller than $\deg\ f(x)$. Then, by the induction hypothesis, there exist polynomials q_1 and r_1 such that

$$f_1(x) = q_1(x)g(x) + r_1(x)$$

and either $r_1 = 0$ or $\deg\ r_1 < \deg\ g$. It is easy to see that, in this case, the polynomials $q(x) = q_1(x) + q^*(x)$ and $r(x) = r_1(x)$ are as required.

It remains to show that the polynomials q and r are unique. Let Q and R be polynomials, possibly different from q and r, satisfying the assertions of the lemma. That is, $f(x) = Q(x)g(x) + R(x)$, and so $(q(x) - Q(x))g(x) = R(x) - r(x)$. If the polynomial on the left-hand side is non-zero, then its degree is at least $\deg\ g$, while the degree of the polynomial on the right-hand side is smaller than $\deg\ g$. This, however, is not possible. ■

Let R be a commutative ring with a multiplicative identity and without zero divisors, and set $R^* := R \setminus \{0\}$. The ring R is said to be a ***Euclidean ring*** if there is a function $\phi : R^* \to \mathbb{N}$ such that $\phi(ab) \geq \phi(a)\phi(b)$, for all $a,\ b \in R^*$; and, further, if $a \in R$, $b \in R^*$, then there are elements $q,\ r \in R$ such that $a = qb + r$, and if $r \neq 0$, then $\phi(r) < \phi(b)$. The previous lemma shows that $\mathbb{F}[x]$ is a Euclidean ring where the rôle of the function ϕ is played by the degree function.

The concept of ***divisibility*** in $\mathbb{F}[x]$ can be defined similarly to the definition of the corresponding concept in the ring of integers. A polynomial $g(x)$ is said to be a ***divisor*** of a polynomial $f(x)$ (the notation is $g \mid f$), if there is a polynomial $q(x) \in \mathbb{F}[x]$ such that $f(x) = q(x)g(x)$. The non-zero elements of $\mathbb{F}$, which are clearly divisors of each of the polynomials, are called the ***trivial divisors*** or ***units***. A non-zero polynomial $f(x) \in \mathbb{F}[x]$ is said

to be ***irreducible***, if whenever $f(x) = q(x)g(x)$ with $q(x),\ g(x) \in \mathbb{F}[x]$, then either q or g is a unit.

Two polynomials $f,\ g \in \mathbb{F}[x]$ are called ***associates***, if there is some $u \in \mathbb{F}^*$ such that $f(x) = ug(x)$.

Using Lemma 1.3, one can easily prove the unique factorisation theorem in the ring of polynomials following the argument of the proof of the corresponding theorem in the ring of integers (see Section 33.1). The rôle of the absolute value of integers is played by the degree of polynomials. Reference to NA!

Theorem 1.4 *An arbitrary polynomial* $0 \neq f \in \mathbb{F}[x]$ *can be written in the form*

$$f(x) = uq_1(x)^{e_1} \cdots q_r(x)^{e_r},$$

where $u \in \mathbb{F}^*$ *is a unit, the polynomials* $q_i \in \mathbb{F}[x]$ *are pairwise non-associate and irreducible, and, further, the numbers* e_i *are positive integers. Furthermore, this decomposition is essentially unique in the sense that whenever*

$$f(x) = UQ_1(x)^{d_1} \cdots Q_s(x)^{d_s}$$

is another such decomposition, then $r = s$*, and, after possibly reordering the factors* Q_i*, the polynomials* q_i *and* Q_i *are associates, and moreover* $d_i = e_i$ *for all* $1 \leq i \leq r$*.*

Two polynomials are said to be ***relatively prime,*** if they have no common irreducible divisors.

A scalar $a \in \mathbb{F}$ is a ***root*** of a polynomial $f \in \mathbb{F}[x]$, if $f(a) = 0$. Here the value $f(a)$ is obtained by substituting a into the place of x in $f(x)$.

Lemma 1.5 *Suppose that* $a \in \mathbb{F}$ *is a root of a polynomial* $f(x) \in \mathbb{F}[x]$*. Then there exists a polynomial* $g(x) \in \mathbb{F}[x]$ *such that* $f(x) = (x - a)g(x)$*. Hence the polynomial* f *may have at most* deg f *roots.*

Proof. By Lemma 1.3, there exists $g(x) \in \mathbb{F}[x]$ and $r \in \mathbb{F}$ such that $f(x) = (x - a)g(x) + r$. Substituting a for x, we find that $r = 0$. The second assertion now follows by induction on deg f from the fact that the roots of g are also roots of f. ∎

The cost of the operations with polynomials

Suppose that $f(x),\ g(x) \in \mathbb{F}[x]$ are polynomials of degree at most n. Then the polynomials $f(x) \pm g(x)$ can obviously be computed using $O(n)$ field operations. The product $f(x)g(x)$ can be obtained, using its definition, by $O(n^2)$ field operations. If the Fast Fourier Transform can be performed over $\mathbb{F}$, then the multiplication can be computed using only $O(n \lg n)$ field operations (see Theorem 32.2). For general fields, the cost of the fastest known multiplication algorithms for polynomials (for instance the Schönhage-Strassen-method) is $O(n \lg n \lg \lg n)$, that is, $\widetilde{O}(n)$ field operations. Reference to NA!

The division with remainder, that is, determining the polynomials $q(x)$ and $r(x)$ for which $f(x) = q(x)g(x) + r(x)$ and either $r(x) = 0$ or deg $r(x) <$ deg $g(x)$, can be performed using $O(n^2)$ field operations following the straightforward method outlined in the proof of Lemma 1.3. There is, however, an algorithm (the Sieveking-Kung algorithm) for the same problem using only $\widetilde{O}(n)$ steps. The details of this algorithm are, however, not discussed here.

Congruence, residue class ring

Let $f(x) \in \mathbb{F}[x]$ with $\deg f = n > 0$, and let $g,\ h \in \mathbb{F}[x]$. We say that g is ***congruent*** to h modulo f, or simply $g \equiv h \pmod{f}$, if f divides the polynomial $g - h$. This concept of congruence is similar to the corresponding concept introduced in the ring of integers (see 33.3.2). It is easy to see from the definition that the relation $\equiv$ is an equivalence relation on the set $\mathbb{F}[x]$. Let $[g]_f$ (or simply $[g]$ if f is clear from the context) denote the equivalence class containing g. From Lemma 1.3 we obtain immediately, for each g, that there is a unique $r \in \mathbb{F}[x]$ such that $[g] = [r]$, and either $r = 0$ (if f divides g) or $\deg r < n$. This polynomial r is called the ***representative*** of the class $[g]$. The set of equivalence classes is traditionally denoted by $\mathbb{F}[x]/(f)$.

Reference to NA!

Lemma 1.6 *Let $f,\ f_1,\ f_2,\ g_1,\ g_2 \in \mathbb{F}[x]$ and let $a \in \mathbb{F}$. Suppose that $f_1 \equiv f_2 \pmod{f}$ and $g_1 \equiv g_2 \pmod{f}$. Then*

$$f_1 + g_1 \equiv f_2 + g_2 \pmod{f},$$

$$f_1 g_1 \equiv f_2 g_2 \pmod{f},$$

and

$$af_1 \equiv af_2 \pmod{f}.$$

Proof. The first congruence is valid, as

$$(f_1 + g_1) - (f_2 + g_2) = (f_1 - f_2) + (g_1 - g_2),$$

and the right-hand side of this is clearly divisible by f. The second and the third congruences follow similarly from the identities

$$f_1 g_1 - f_2 g_2 = (f_1 - f_2)g_1 + (g_1 - g_2)f_2$$

and

$$af_1 - af_2 = a(f_1 - f_2),$$

respectively. ■

The previous lemma makes it possible to define the sum and the product of two congruence classes $[g]_f$ and $[h]_f$ as $[g]_f + [h]_f := [g + h]_f$ and $[g]_f[h]_f := [gh]_f$, respectively. The lemma claims that the sum and the product are independent of the choice of the congruence class representatives. The same way, we may define the action of $\mathbb{F}$ on the set of congruence classes: we set $a[g]_f := [ag]_f$.

Theorem 1.7 *Suppose that $f(x) \in \mathbb{F}[x]$ and that* $\deg f = n > 0$.
(i) The set of residue classes $\mathbb{F}[x]/(f)$ is a commutative ring with an identity under the operations $+$ and $\cdot$ defined above.
(ii) The ring $\mathbb{F}[x]/(f)$ contains the field $\mathbb{F}$ as a subring, and it is an n-dimensional vector space over $\mathbb{F}$. Further, the residue classes $[1],\ [x], \ldots, [x^{n-1}]$ form a basis of $\mathbb{F}[x]/(f)$.
(iii) If f is an irreducible polynomial in $\mathbb{F}[x]$, then $\mathbb{F}[x]/(f)$ is a field.

Proof. (i) The fact that $\mathbb{F}[x]/(f)$ is a ring follows easily from the fact that $\mathbb{F}[x]$ is a ring. Let us, for instance, verify the distributive property:

$$[g]([h_1]+[h_2]) = [g][h_1+h_2] = [g(h_1+h_2)] = [gh_1+gh_2] = [gh_1]+[gh_2] = [g][h_1]+[g][h_2].$$

The zero element of $\mathbb{F}[x]/(f)$ is the class $[0]$, the additive inverse of the class $[g]$ is the class $[-g]$, while the multiplicative identity element is the class $[1]$. The details are left to the reader.

(ii) The set $\{[a] \mid a \in \mathbb{F}\}$ is a subring isomorphic to $\mathbb{F}$. The correspondence is obvious: $a \leftrightarrow [a]$. By part (i), $\mathbb{F}[x]/(f)$ is an additive Abelian group, and the action of $\mathbb{F}$ satisfies the vector space axioms. This follows from the fact that the polynomial ring is itself a vector space over $\mathbb{F}$. Let us, for example, verify the distributive property:

$$a([h_1] + [h_2]) = a[h_1 + h_2] = [a(h_1 + h_2)] = [ah_1 + ah_2] = [ah_1] + [ah_2] = a[h_1] + a[h_2] \ .$$

The other properties are left to the reader.

We claim that the classes $[1],\ [x], \dots, [x^{n-1}]$ are linearly independent. For, if

$$[0] = a_0[1] + a_1[x] + \cdots + a_{n-1}[x^{n-1}] = [a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}] \ ,$$

then $a_0 = \cdots = a_{n-1} = 0$, as the zero polynomial is the unique polynomial with degree less than n that is divisible by f. On the other hand, for a polynomial g, the degree of the class representative of $[g]$ is less than n. Thus the class $[g]$ can be expressed as a linear combination of the classes $[1], [x], \dots, [x^{n-1}]$. Hence the classes $[1], [x], \dots, [x^{n-1}]$ form a basis of $\mathbb{F}[x]/(f)$, and so $\dim_{\mathbb{F}} \mathbb{F}[x]/(f) = n$.

(iii) Suppose that f is irreducible. First we show that $\mathbb{F}[x]/(f)$ has no zero divisors. If $[0] = [g][h] = [gh]$, then f divides gh, and so f divides either g or h. That is, either $[g] = 0$ or $[h] = 0$. Suppose now that $g \in \mathbb{F}[x]$ with $[g] \neq [0]$. We claim that the classes $[g][1], [g][x], \dots, [g][x^{n-1}]$ are linearly independent. Indeed, an equation $[0] = a_0[g][1] + \cdots + a_{n-1}[g][x^{n-1}]$ implies $[0] = [g][a_0 + \cdots + a_{n-1}x^{n-1}]$, and, in turn, it also yields that $a_0 = \cdots = a_{n-1} = 0$. Therefore the classes $[g][1], [g][x], \dots, [g][x^{n-1}]$ form a basis of $\mathbb{F}[x]/(f)$. Hence there exist coefficients $b_i \in \mathbb{F}$ for which

$$[1] = b_0[g][1] + \cdots + b_{n-1}[g][x^{n-1}] = [g][b_0 + \cdots + b_{n-1}x^{n-1}] \ .$$

Thus we find that the class $[0] \neq [g]$ has a multiplicative inverse, and so $\mathbb{F}[x]/(f)$ is a field, as required. ∎

We note that the converse of part (iii) of the previous theorem is also true, and its proof is left to the reader (Exercise 1.1-1.).

1.2. Example. We usually represent the elements of the residue class ring $\mathbb{F}[x]/(f)$ by their representatives, which are polynomials with degree less than $\deg f$.

1. Suppose that $\mathbb{F} = \mathbb{F}_2$ is the field of two elements, and let $f(x) = x^3 + x + 1$. Then the ring $\mathbb{F}[x]/(f)$ has 8 elements, namely

$$[0],\ [1],\ [x],\ [x+1],\ [x^2],\ [x^2+1],\ [x^2+x],\ [x^2+x+1].$$

Practically speaking, the addition between the classes is the is addition of polynomials. For instance

$$[x^2 + 1] + [x^2 + x] = [x + 1] \ .$$

When computing the product, we compute the product of the representatives, and substitute it (or reduce it) with its remainder after dividing by f. For instance,

$$[x^2 + 1] \cdot [x^2 + x] = [x^4 + x^3 + x^2 + x] = [x + 1] \ .$$

The polynomial f is irreducible over $\mathbb{F}_2$, since it has degree 3, and has no roots. Hence the residue class ring $\mathbb{F}[x]/(f)$ is a field.

2. Let $\mathbb{F} = \mathbb{R}$ and let $f(x) = x^2 - 1$. The elements of the residue class ring are the classes of the form $[ax + b]$ where $a,\ b \in \mathbb{R}$. The ring $\mathbb{F}[x]/(f)$ is not a field, since f is not irreducible. For instance, $[x + 1][x - 1] = [0]$.

Lemma 1.8 *Let $\mathbb{L}$ be a field containing a field $\mathbb{F}$ and let $\alpha \in \mathbb{L}$.*
(i) If $\mathbb{L}$ is finite-dimensional as a vector space over $\mathbb{F}$, then there is a non-zero polynomial $f \in \mathbb{F}[x]$ such that α is a root of f.
(ii) Assume that there is a polynomial $f \in \mathbb{F}[x]$ with $f(\alpha) = 0$, and let g be such a polynomial with minimal degree. Then the polynomial g is irreducible in $\mathbb{F}[x]$. Further, if $h \in \mathbb{F}[x]$ with $h(\alpha) = 0$ then g is a divisor of h.

Proof. (i) For a sufficiently large n, the elements $1, \alpha, \ldots, \alpha^n$ are linearly dependent over $\mathbb{F}$. A linear dependence gives a polynomial $0 \neq f \in \mathbb{F}[x]$ such that $f(\alpha) = 0$.

(ii) If $g = g_1 g_2$, then, as $0 = g(\alpha) = g_1(\alpha)g_2(\alpha)$, the element α is a root of either g_1 or g_2. As g was chosen to have minimal degree, one of the polynomials $g_1,\ g_2$ is a unit, and so g is irreducible. Finally, let $h \in \mathbb{F}[x]$ such that $h(\alpha) = 0$. Let $q,\ r \in \mathbb{F}[x]$ be the polynomials as in Lemma 1.3 for which $h(x) = q(x)g(x) + r(x)$. Substituting α for x into the last equation, we obtain $r(\alpha) = 0$, which is only possible if $r = 0$. ∎

Definition 1.9 *The polynomial $g \in \mathbb{F}[x]$ in the last lemma is said to be a **minimal polynomial** of α.*

It follows from the previous lemma that the minimal polynomial is unique up to a scalar multiple. It will often be helpful to assume that the leading coefficient (the coefficient of the term with the highest degree) of the minimal polynomial g is 1.

Corollary 1.10 *Let $\mathbb{L}$ be a field containing $\mathbb{F}$, and let $\alpha \in \mathbb{L}$. Suppose that $f \in \mathbb{F}[x]$ is irreducible and that $f(\alpha) = 0$. Then f is a minimal polynomial of α.*

Proof. Suppose that g is a minimal polynomial of α. By the previous lemma, $g \mid f$ and g is irreducible. This is only possible if the polynomials f and g are associates. ∎

Let $\mathbb{L}$ be a field containing $\mathbb{F}$ and let $\alpha \in \mathbb{L}$. Let $\mathbb{F}(\alpha)$ denote the smallest subfield of $\mathbb{L}$ that contains $\mathbb{F}$ and α.

Theorem 1.11 *Let $\mathbb{L}$ be a field containing $\mathbb{F}$ and let $\alpha \in \mathbb{L}$. Suppose that $f \in \mathbb{F}[x]$ is a minimal polynomial of α. Then the field $\mathbb{F}(\alpha)$ is isomorphic to the field $\mathbb{F}[x]/(f)$. More precisely, there exists an isomorphism $\phi : \mathbb{F}[x]/(f) \to \mathbb{F}(\alpha)$ such that $\phi(a) = a$, for all $a \in \mathbb{F}$, and $\phi([x]_f) = \alpha$. The map ϕ is also an isomorphism of vector spaces over $\mathbb{F}$, and so* $\dim_{\mathbb{F}} \mathbb{F}(\alpha) = \deg f$.

Proof. Let us consider the map $\psi : \mathbb{F}[x] \to \mathbb{L}$, which maps a polynomial $g \in \mathbb{F}[x]$ into $g(\alpha)$. This is clearly a ring homomorphism, and $\psi(\mathbb{F}[x]) \subseteq \mathbb{F}(\alpha)$. We claim that $\psi(g) = \psi(h)$ if and only if $[g]_f = [h]_f$. Indeed, $\psi(g) = \psi(h)$ holds if and only if $\psi(g - h) = 0$, that is, if $g(\alpha) - h(\alpha) = 0$, which, by Lemma 1.8, is equivalent to $f \mid g - h$, and this amounts to saying that $[g]_f = [h]_f$. Suppose that ϕ is the map $\mathbb{F}[x]/(f) \to \mathbb{F}(\alpha)$ induced by ψ, that is, $\phi([g]_f) := \psi(g)$. By the argument above, the map ϕ is one-to-one. Routine computation

shows that ϕ is a ring, and also a vector space, homomorphism. As $\mathbb{F}[x]/(f)$ is a field, its homomorphic image $\phi(\mathbb{F}[x]/(f))$ is also a field. The field $\phi(\mathbb{F}[x]/(f))$ contains $\mathbb{F}$ and α, and so necessarily $\phi(\mathbb{F}[x]/(f)) = \mathbb{F}(\alpha)$. ■

Euclidean algorithm and the greatest common divisor

Let $f(x),\ g(x) \in \mathbb{F}[x]$ be polynomials such that $g(x) \neq 0$. Set $f_0 = f$, $f_1 = g$ and define the polynomials q_i and f_i using division with reminder as follows:

$$f_0(x) = q_1(x)f_1(x) + f_2(x)\ ,$$

$$f_1(x) = q_2(x)f_2(x) + f_3(x)\ ,$$

$$\vdots$$

$$f_{k-2}(x) = q_{k-1}(x)f_{k-1}(x) + f_k(x)\ ,$$

$$f_{k-1}(x) = q_k(x)f_k(x) + f_{k+1}(x)\ .$$

Note that if $1 < i < k$ then $\deg f_{i+1}$ is smaller than $\deg f_i$. We form this sequence of polynomials until we obtain that $f_{k+1} = 0$. By Lemma 1.3, this defines a finite process. Let n be the maximum of $\deg f$ and $\deg g$. As, in each step, we decrease the degree of the polynomials, we have $k \leq n + 1$. The computation outlined above is usually referred to as the ***Euclidean algorithm***. A version of this algorithm for the ring of integers is described in Section 33.2. Reference to NA!

We say that the polynomial $h(x)$ is the ***greatest common divisor*** of the polynomials $f(x)$ and $g(x)$, if $h(x) \mid f(x)$, $h(x) \mid g(x)$, and, if a polynomial $h_1(x)$ is a divisor of f and g, then $h_1(x)$ is a divisor of $h(x)$. The usual notation for the greatest common divisor of $f(x)$ and $g(x)$ is $\gcd(f(x), g(x))$. It follows from Theorem 1.4 that $\gcd(f(x), g(x))$ exists and it is unique up to a scalar multiple.

Theorem 1.12 *Suppose that $f(x),\ g(x) \in \mathbb{F}[x]$ are polynomials, that $g(x) \neq 0$, and let n be the maximum of* $\deg f$ *and* $\deg g$. *Assume, further, that the number k and the polynomial f_k are defined by the procedure above. Then*
(i) $\gcd(f(x), g(x)) = f_k(x)$.
(ii) There are polynomials $F(x),\ G(x)$ with degree at most n such that

$$f_k(x) = F(x)f(x) + G(x)g(x)\ . \tag{1.1}$$

(iii) With a given input $f,\ g$, the polynomials $F(x),\ G(x),\ f_k(x)$ can be computed using $O(n^3)$ field operations in $\mathbb{F}$.

Proof. (i) Going backwards in the Euclidean algorithm, it is easy to see that the polynomial f_k divides each of the f_i, and so it divides both f and g. The same way, if a polynomial $h(x)$ divides f and g, then it divides f_i, for all i, and, in particular, it divides f_k. Thus $\gcd(f(x), g(x)) = f_k(x)$.

(ii) The claim is obvious if $f = 0$, and so we may assume without loss of generality that $f \neq 0$. Starting at the beginning of the Euclidean sequence, it is easy to see that there are polynomials $F_i(x),\ G_i(x) \in \mathbb{F}[x]$ such that

$$F_i(x)f(x) + G_i(x)g(x) = f_i(x). \tag{1.2}$$

We observe that (1.2) also holds if we substitute $F_i(x)$ by its remainder $F_i^*(x)$ after dividing by g and substitute $G_i(x)$ by its remainder $G_i^*(x)$ after dividing by f. In order to see this, we compute

$$F_i^*(x)f(x) + G_i^*(x)g(x) \equiv f_i(x) \pmod{f(x)g(x)},$$

and notice that the degree of the polynomials on both sides of this congruence is smaller than $(\deg f)(\deg g)$. This gives

$$F_i^*(x)f(x) + G_i^*(x)g(x) = f_i(x).$$

(iii) Once we determined the polynomials f_{i-1}, f_i, F_i^* and G_i^*, the polynomials f_{i+1}, F_{i+1}^* and G_{i+1}^* can be obtained using $O(n^2)$ field operations in $\mathbb{F}$. Initially we have $F_1^* = 1$ and $G_2^* = -q_1$. As $k \leq n+1$, the claim follows. ∎

Remark. Traditionally, the Euclidean algorithm is only used to compute the greatest common divisor. The version that also computes the polynomials $F(x)$ and $G(x)$ in (1.1) is usually called the extended Euclidean algorithm. In Chapter **??** the reader can find a discussion of the Euclidean algorithm for polynomials. It is relatively easy to see that the polynomials $f_k(x)$, $F(x)$, and $G(x)$ in (1.1) can, in fact, be computed using $O(n^2)$ field operations. The cost of the asymptotically best method is $\widetilde{O}(n)$.

The derivative of a polynomial is often useful when investigating multiple factors. The ***derivative*** of the polynomial

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \in \mathbb{F}[x]$$

is the polynomial

$$f'(x) = a_1 + 2a_2x + \cdots + na_nx^{n-1}\ .$$

It follows immediately from the definition that the map $f(x) \mapsto f'(x)$ is an $\mathbb{F}$-linear mapping $\mathbb{F}[x] \to \mathbb{F}[x]$. Further, for $f(x),\ g(x) \in \mathbb{F}[x]$ and $a \in \mathbb{F}$, the equations $(f(x)+g(x))' = f'(x)+g'(x)$ and $(af(x))' = af'(x)$ hold. The derivative of a product can be computed using the ***Leibniz rule***: for all $f(x),\ g\ (x) \in \mathbb{F}[x]$ we have $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$. As the derivation is a linear map, in order to show that the Leibniz rule is valid, it is enough to verify it for polynomials of the form $f(x) = x^i$ and $g(x) = x^j$. It is easy to see that, for such polynomials, the Leibniz rule is valid.

The derivative $f'(x)$ is sensitive to multiple factors in the irreducible factorisation of $f(x)$.

Lemma 1.13 *Let $\mathbb{F}$ be an arbitrary field, and assume that $f(x) \in \mathbb{F}[x]$ and $f(x) = u^k(x)v(x)$ where $u(x),\ v(x) \in \mathbb{F}[x]$. Then $u^{k-1}(x)$ divides the derivative $f'(x)$ of the polynomial $f(x)$.*

Proof. Using induction on k and the Leibniz rule, we find $(u^k(x))' = ku^{k-1}(x)u'(x)$. Thus, applying the Leibniz rule again, $f'(x) = u^{k-1}(x)(ku'(x)v(x) + u^k(x)v'(x))$. Hence $u^{k-1}(x) \mid f'(x)$. ∎

In many cases the converse of the last lemma also holds.

Lemma 1.14 *Let $\mathbb{F}$ be an arbitrary field, and assume that $f(x) \in \mathbb{F}[x]$ and $f(x) = u(x)v(x)$ where the polynomials $u(x)$ and $v(x)$ are relatively prime. Suppose further that $u'(x) \neq 0$ (for instance $\mathbb{F}$ has characteristic 0 and $u(x)$ is non-constant). Then the derivative $f'(x)$ is not divisible by $u(x)$.*

Proof. By the Leibniz rule, $f'(x) = u(x)v'(x) + u'(x)v(x) \equiv u'(x)v(x) \pmod{u(x)}$. Since $\deg u'(x)$ is smaller than $\deg u(x)$, we obtain that $u'(x)$ is not divisible by $u(x)$, and neither is the product $u'(x)v(x)$, as $u(x)$ and $v(x)$ are relatively prime. ■

The Chinese remainder theorem for polynomials

Using the following theorem, the ring $\mathbb{F}[x]/(f)$ can be assembled from rings of the form $\mathbb{F}[x]/(g)$ where $g \mid f$.

Theorem 1.15 *(Chinese remainder theorem for polynomials) Let $f_1, \ldots, f_k \in \mathbb{F}[x]$ pairwise relatively prime polynomials with positive degree and set $f = f_1 \cdots f_k$. Then the rings $\mathbb{F}[x]/(f)$ and $\mathbb{F}[x]/(f_1) \oplus \cdots \oplus \mathbb{F}[x]/(f_k)$ are isomorphic. The mapping realizing the isomorphism is*

$$\phi : [g]_f \mapsto ([g]_{f_1}, \ldots, [g]_{f_k}), \quad g \in \mathbb{F}[x] \, .$$

Proof. First we note that the map ϕ is well-defined. If $h \in [g]_f$, then $h = g + f^* f$, which implies that h and g give the same remainder after division by the polynomial f_i, that is, $[h]_{f_i} = [g]_{f_i}$.

The mapping ϕ is clearly a ring homomorphism, and it is also a linear mapping between two vector spaces over $\mathbb{F}$. The mapping ϕ is one-to-one; for, if $\phi([g]) = \phi([h])$, then $\phi([g - h]) = (0, \ldots, 0)$, that is, $f_i \mid g - h$ $(1 \le i \le k)$, which gives $f \mid g - h$ and $[g] = [h]$.

The dimensions of the vector spaces $\mathbb{F}[x]/(f)$ and $\mathbb{F}[x]/(f_1) \oplus \cdots \oplus \mathbb{F}[x]/(f_k)$ coincide: indeed, both spaces have dimension $\deg f$. Lemma 1.1 implies that ϕ is an isomorphism between vector spaces. It only remains to show that ϕ^{-1} preserves the multiplication; this, however, is left to the reader. ■

Exercises

1.1-1 Let $f \in \mathbb{F}[x]$ be polynomial. Show that the residue class ring $\mathbb{F}[x]/(f)$ has no zero divisors if and only if f is irreducible.

1.1-2 Let R be a commutative ring with an identity. A subset $I \subseteq R$ is said to be an ***ideal***, if I is an additive subgroup, and $a \in I$, $b \in R$ imply $ab \in I$. Show that R is a field if and only if its ideals are exactly $\{0\}$ and R.

1.1-3 Let $a_1, \ldots, a_k \in R$. Let $(a_1, \ldots, a_k)$ denote the smallest ideal in R that contains the elements a_i. Show that $(a_1, \ldots, a_k)$ always exists, and it consists of the elements of the form $b_1 a_1 + b_2 a_2 + \cdots + b_k a_k$ where $b_1, \ldots, b_k \in R$.

1.1-4 A commutative ring R with an identity and without zero divisors is said to be a ***principal ideal domain*** if, for each ideal I of R, there is an element $a \in I$ such that (using the notation of the previous exercise) $I = (a)$. Show that $\mathbf{Z}$ and $\mathbb{F}[x]$ where $\mathbb{F}$ is a field, are principal ideal domains.

1.1-5 Suppose that S is a commutative ring with an identity, that I an ideal in S, and that $a, b \in S$. Define a relation on S as follows: $a \equiv b \pmod{I}$ if and only if $a - b \in I$. Verify the following:

a.) The relation $\equiv$ is an equivalence relation on S.

b.) Let $[a]_I$ denote the equivalence class containing an element a, and let S/I denote the set of equivalence classes. Set $[a]_I + [b]_I := [a + b]_I$, and $[a]_I [b]_I := [ab]_I$. Show that, with respect to these operations, S/I is a commutative ring with an identity. *Hint:* Follow the argument in the proof of Theorem 1.7.

1.1-6 Let $\mathbb{F}$ be a field and let $f(x), g(x) \in \mathbb{F}[x]$ such that $\gcd(f(x), g(x)) = 1$. Show that there exists a polynomial $h(x) \in \mathbb{F}[x]$ such that $h(x)g(x) \equiv 1 \pmod{f(x)}$. *Hint:* Use the

Euclidean algorithm.

1.2. Finite fields

Finite fields, that is, fields with a finite number of elements, play an important rôle in mathematics and in several of its application areas, for instance, in computing. They are also fundamental in many important constructions. In this section we summarise the most important results in the theory of finite fields, putting an emphasis on the problem of their construction.

In this section p denotes a prime number, and q denotes a power of p with a positive integer exponent.

Theorem 1.16 *Suppose that $\mathbb{F}$ is a finite field. Then there is a prime number p such that the prime field of $\mathbb{F}$ is isomorphic to $\mathbb{F}_p$ (the field of residue classes modulo p). Further, the field $\mathbb{F}$ is a finite dimensional vector space over $\mathbb{F}_p$, and the number of its elements is a power of p. In fact, if $\dim_{\mathbb{F}_p} \mathbb{F} = d$, then $|\mathbb{F}| = p^d$.*

Proof. The characteristic of $\mathbb{F}$ must be a prime, say p, as a field with characteristic zero must have infinitely many elements. Thus the prime field P of $\mathbb{F}$ is isomorphic to $\mathbb{F}_p$. Since P is a subfield, the field $\mathbb{F}$ is a vector space over P. Let $\alpha_1, \ldots, \alpha_d$ be a basis of $\mathbb{F}$ over P. Then each $\alpha \in \mathbb{F}$ can be written uniquely in the form $\sum_{j=1}^{d} a_i \alpha_i$ where $a_i \in P$. Hence $|\mathbb{F}| = p^d$. ■

In a field $\mathbb{F}$, the set of non-zero elements (the multiplicative group of $\mathbb{F}$) is denoted by $\mathbb{F}^*$. From Theorem 1.2 we immediately obtain the following result.

Theorem 1.17 *If $\mathbb{F}$ is a finite field, then its multiplicative group $\mathbb{F}^*$ is cyclic.*

A generator of the group $\mathbb{F}^*$ is said to be a ***primitive element***. If $|\mathbb{F}| = q$ and α is a primitive element of $\mathbb{F}$, then the elements of $\mathbb{F}$ are $0, \alpha, \alpha^2, \ldots, \alpha^{q-1} = 1$.

Corollary 1.18 *Suppose that $\mathbb{F}$ is a finite field with order p^d and let α be a primitive element of $\mathbb{F}$. Let $g \in \mathbb{F}_p[x]$ be a minimal polynomial of α over $\mathbb{F}_p$. Then g is irreducible in $\mathbb{F}_p[x]$, the degree of g is d, and $\mathbb{F}$ is isomorphic to the field $\mathbb{F}_p[x]/(g)$.*

Proof. Since the element α is primitive in $\mathbb{F}$, we have $\mathbb{F} = \mathbb{F}_p(\alpha)$. The rest of the lemma follows from Lemma 1.8 and from Theorem 1.11. ■

Theorem 1.19 *Let $\mathbb{F}$ be a finite field with order q. Then*
(i) (Fermat's little theorem) *If $\beta \in \mathbb{F}^*$, then $\beta^{q-1} = 1$.*
(ii) If $\beta \in \mathbb{F}$, then $\beta^q = \beta$.

Proof. (i) Suppose that $\alpha \in \mathbb{F}^*$ is a primitive element. Then we may choose an integer i such that $\beta = \alpha^i$. Therefore

$$\beta^{q-1} = (\alpha^i)^{q-1} = (\alpha^{q-1})^i = 1^i = 1.$$

(ii) Clearly, if $\beta = 0$ then this claim is true, while, for $\beta \neq 0$, the claim follows from part (i). ■

Theorem 1.20 *Let $\mathbb{F}$ be a field with q elements. Then*

$$x^q - x = \prod_{\alpha \in \mathbb{F}} (x - \alpha) \ .$$

Proof. By Theorem 1.19 and Lemma 1.5, the product on the right-hand side is a divisor of the polynomial $x^q - x \in \mathbb{F}[x]$. Now the assertion follows, as the degrees and the leading coefficients of the two polynomials in the equation coincide. ∎

Corollary 1.21 *Arbitrary two finite fields with the same number of elements are isomorphic.*

Proof. Suppose that $q = p^d$, and that both $\mathbb{K}$ and $\mathbb{L}$ are fields with q elements. Let β be a primitive element in $\mathbb{L}$. Then Corollary 1.18 implies that a minimal polynomial $g(x) \in \mathbb{F}_p[x]$ of β over $\mathbb{F}_p$ is irreducible (in $\mathbb{F}_p[x]$) with degree d. Further, $\mathbb{L} \cong \mathbb{F}_p[x]/(g(x))$. By Lemma 1.8 and Theorem 1.19, the minimal polynomial g is a divisor of the polynomial $x^q - x$. Applying Theorem 1.20 to $\mathbb{K}$, we find that the polynomial $x^q - x$, and also its divisor $g(x)$, can be factored as a product of linear terms in $\mathbb{K}[x]$, and so $g(x)$ has at least one root α in $\mathbb{K}$. As $g(x)$ is irreducible in $\mathbb{F}_p[x]$, it must be a minimal polynomial of α (see Corollary 1.10), and so $\mathbb{F}_p(\alpha)$ is isomorphic to the field $\mathbb{F}_p[x]/(g(x))$. Comparing the number of elements in $\mathbb{F}_p(\alpha)$ and in $\mathbb{K}$, we find that $\mathbb{F}_p(\alpha) = \mathbb{K}$, and further, that $\mathbb{K}$ and $\mathbb{L}$ are isomorphic. ∎

In the sequel, we let $\mathbb{F}_q$ denote the field with q elements, provided it exists. In order to prove the existence of such a field for each prime-power q, the following two facts will be useful.

Lemma 1.22 *If p is a prime number and j is an integer such that $0 < j < p$, then $p \mid \binom{p}{j}$.*

Proof. On the one hand, the number $\binom{p}{j}$ is an integer. On the other hand, $\binom{p}{j} = p(p-1)\cdots(p-j+1)/j!$ is a fraction such that, for $0 < j < p$, its numerator is divisible by p, but its denominator is not. ∎

Lemma 1.23 *Let R be a commutative ring and let p be a prime such that $pr = 0$ for all $r \in R$. Then the map $\Phi_p : R \to R$ mapping $r \mapsto r^p$ is a ring homomorphism.*

Proof. Suppose that $r,\ s \in R$. Clearly,

$$\Phi_p(rs) = (rs)^p = r^p s^p = \Phi_p(r)\Phi_p(s) \ .$$

By the previous lemma,

$$\Phi_p(r+s) = (r+s)^p = \sum_{j=0}^{p} \binom{p}{j} r^{p-j} s^j = r^p + s^p = \Phi_p(r) + \Phi_p(s) \ .$$

We obtain in the same way that $\Phi_p(r-s) = \Phi_p(r) - \Phi_p(s)$. ∎

The homomorphism Φ_p in the previous lemma is called the ***Frobenius endomorphism***.

Theorem 1.24 *Assume that the polynomial $g(x) \in \mathbb{F}_q[x]$ is irreducible, and, for a positive integer d, it is a divisor of the polynomial $x^{q^d} - x$. Then the degree of $g(x)$ divides d.*

Proof. Let n be the degree of $g(x)$, and suppose, by contradiction, that $d = tn + s$ where $0 < s < n$. The assumption that $g(x) \mid x^{q^d} - x$ can be rephrased as $x^{q^d} \equiv x \pmod{g(x)}$. However, this means that, for an arbitrary polynomial $u(x) = \sum_{i=0}^{N} u_i x^i \in \mathbb{F}_q[x]$, we have

$$u(x)^{q^d} = \sum_{i=0}^{N} u_i^{q^d} x^{iq^d} = \sum_{i=0}^{N} u_i (x^{q^d})^i \equiv \sum_{i=0}^{N} u_i x^i = u(x) \pmod{g(x)}.$$

Note that we applied Lemma 1.23 to the ring $R = \mathbb{F}_q[x]/(g(x))$, and Theorem 1.19 to $\mathbb{F}_q$. The residue class ring $\mathbb{F}_q[x]/(g(x))$ is isomorphic to the field $\mathbb{F}_{q^n}$, which has q^n elements. Let $u(x) \in \mathbb{F}_q[x]$ be a polynomial for which $u(x) \pmod{g(x)}$ is a primitive element in the field $\mathbb{F}_{q^n}$. That is, $u(x)^{q^n-1} \equiv 1 \pmod{g(x)}$, but $u(x)^j \not\equiv 1 \pmod{g(x)}$ for $j = 1, \ldots, q^n - 2$. Therefore,

$$u(x) \equiv u(x)^{q^d} = u(x)^{q^{tn+s}} = (u(x)^{q^{nt}})^{q^s} \equiv u(x)^{q^s} \pmod{g(x)},$$

and so $u(x)(u(x)^{q^s-1} - 1) \equiv 0 \pmod{g(x)}$. Since the residue class ring $\mathbb{F}_q[x]/(g(x))$ is a field, $u(x) \not\equiv 0 \pmod{g(x)}$, but we must have $u(x)^{q^s-1} \equiv 1 \pmod{g(x)}$. As $0 \le q^s - 1 < q^n - 1$, this contradicts to the primitivity of $u(x) \pmod{g(x)}$. ∎

Theorem 1.25 *For an arbitrary prime p and positive integer d, there exists a field with p^d elements.*

Proof. We use induction on d. The claim clearly holds if $d = 1$. Now let $d > 1$ and let r be a prime divisor of d. By the induction hypothesis, there is a field with $q = p^{(d/r)}$ elements. By Theorem 1.24, each of the irreducible factors, in $\mathbb{F}_q[x]$, of the the polynomial $f(x) = x^{q^r} - x$ has degree either 1 or r. Further, $f'(x) = (x^{q^r} - x)' = -1$, and so, by Lemma 1.13, $f(x)$ is square-free. Over $\mathbb{F}_q$, the number of linear factors of $f(x)$ is at most q, and so is the degree of their product. Hence there exist at least $(q^r - q)/r \ge 1$ polynomials with degree r that are irreducible in $\mathbb{F}_q[x]$. Let $g(x)$ be such a polynomial. Then the field $\mathbb{F}_q[x]/(g(x))$ is isomorphic to the field with $q^r = p^d$ elements. ∎

Corollary 1.26 *For each positive integer d, there is an irreducible polynomial $f \in \mathbb{F}_p[x]$ with degree d.*

Proof. Take a minimal polynomial over $\mathbb{F}_p$ of a primitive element in $\mathbb{F}_{p^d}$. ∎

A little bit later, in Theorem 1.31, we will prove a stronger statement: a random polynomial in $\mathbb{F}_p[x]$ with degree d is irreducible with high probability.

Subfields of finite fields

The following theorem describes all subfields of a finite field.

Theorem 1.27 *The field $\mathbb{F} = \mathbb{F}_{p^n}$ contains a subfield isomorphic to $\mathbb{F}_{p^k}$, if and only if $k \mid n$. In this case, there is exactly one subfield in $\mathbb{F}$ that is isomorphic to $\mathbb{F}_{p^k}$.*

Proof. The condition that $k \mid n$ is necessary, since the larger field is a vector space over the smaller field, and so $p^n = (p^k)^l$ must hold with a suitable integer l.

Conversely, suppose that $k \mid n$, and let $f \in \mathbb{F}_p[x]$ be an irreducible polynomial with degree k. Such a polynomial exists by Corollary 1.26. Let $q = p^k$. Applying Theorem 1.19, we obtain, in $\mathbb{F}_p[x]/(f)$, that $x^q \equiv x \pmod{f}$, which yields $x^{p^n} = x^{q^l} \equiv x \pmod{f}$. Thus f must be a divisor of the polynomial $x^{p^n} - x$. Using Theorem 1.20, we find that f has a root

α in $\mathbb{F}$. Now we may prove in the usual way that the subfield $\mathbb{F}_p(\alpha)$ is isomorphic to $\mathbb{F}_{p^k}$.

The last assertion is valid, as the elements of $\mathbb{F}_q$ are exactly the roots of $x^q - x$ (Theorem 1.20), and this polynomial can have, in an arbitrary field, at most q roots. ■

The structure of irreducible polynomials

Next we prove an important property of the irreducible polynomials over finite fields.

Theorem 1.28 *Assume that $\mathbb{F}_q \subseteq \mathbb{F}$ are finite fields, and let $\alpha \in \mathbb{F}$. Let $f \in \mathbb{F}_q[x]$ be the minimal polynomial of α over $\mathbb{F}_q$ with leading coefficient 1, and suppose that* $\deg f = d$. *Then*

$$f(x) = (x - \alpha)(x - \alpha^q) \cdots (x - \alpha^{q^{d-1}}).$$

Moreover, the elements $\alpha, \alpha^q, \ldots, \alpha^{q^{d-1}}$ are pairwise distinct.

Proof. Let $f(x) = a_0 + a_1 x + \cdots + x^d$. If $\beta \in \mathbb{F}$ with $f(\beta) = 0$, then, using Lemma 1.23 and Theorem 1.19, we obtain

$$0 = f(\beta)^q = (a_0 + a_1\beta + \cdots + \beta^d)^q = a_0^q + a_1^q\beta^q + \cdots + \beta^{dq} = a_0 + a_1\beta^q + \cdots + \beta^{qd} = f(\beta^q).$$

Thus β^q is also a root of f.

As α is a root of f, the argument in the previous paragraph shows that so are the elements $\alpha, \alpha^q, \ldots, \alpha^{q^{d-1}}$. Hence it suffices to show, that they are pairwise distinct. Suppose, by contradiction, that $\alpha^{q^i} = \alpha^{q^j}$ and that $0 \le i < j < d$. Let $\beta = \alpha^{q^i}$ and let $l = j - i$. By assumption, $\beta = \beta^{q^l}$, which, by Lemma 1.8, means that $f(x) \mid x^{q^l} - x$. From Theorem 1.24, we obtain, in this case, that $d \mid l$, which is a contradiction, as $l < d$. ■

This theorem shows that a polynomial f which is irreducible over a finite field cannot have multiple roots. Further, all the roots of f can be obtained from a single root taking q-th powers repeatedly.

Automorphisms

In this section we characterise certain automorphisms of finite fields.

Definition 1.29 *Suppose that $\mathbb{F}_q \subseteq \mathbb{F}$ are finite fields. The map $\Psi : \mathbb{F} \to \mathbb{F}$ is an $\mathbb{F}_q$-**automorphism** of the field $\mathbb{F}$, if it is an isomorphism between rings, and $\Psi(a) = a$ holds for all $a \in \mathbb{F}_q$.*

Recall that the map $\Phi = \Phi_q : \mathbb{F} \to \mathbb{F}$ is defined as follows: $\Phi(\alpha) = \alpha^q$ where $\alpha \in \mathbb{F}$.

Theorem 1.30 *The set of $\mathbb{F}_q$-automorphisms of the field $\mathbb{F} = \mathbb{F}_{q^d}$ is formed by the maps $\Phi, \Phi^2, \ldots, \Phi^d = id$.*

Proof. By Lemma 1.23, the map $\Phi : \mathbb{F} \to \mathbb{F}$ is a ring homomorphism. The map Φ is obviously one-to-one, and hence it is also an isomorphism. It follows from Theorem 1.19, that Φ leaves the elements $\mathbb{F}_q$ fixed. Thus the maps Φ^j are $\mathbb{F}_q$-automorphisms of $\mathbb{F}$.

Suppose that $f(x) = a_0 + a_1 x + \cdots + x^d \in \mathbb{F}_q[x]$, and $\beta \in \mathbb{F}$ with $f(\beta) = 0$, and that Ψ is an $\mathbb{F}_q$-automorphism of $\mathbb{F}$. We claim that $\Psi(\beta)$ is a root of f. Indeed,

$$0 = \Psi(f(\beta)) = \Psi(a_0) + \Psi(a_1)\Psi(\beta) + \cdots + \Psi(\beta)^d = f(\Psi(\beta)) \ .$$

Let β be a primitive element of $\mathbb{F}$ and assume now that $f \in \mathbb{F}_q[x]$ is a minimal polynomial of β. By the observation above and by Theorem 1.28, $\Psi(\beta) = \beta^{q^j}$, with some $0 \le j < d$, that is, $\Psi(\beta) = \Phi^j(\beta)$. Hence the images of a generating element of $\mathbb{F}$ under the automorphisms Ψ and Φ^j coincide, which gives $\Psi = \Phi^j$. ■

The construction of finite fields

Let $q = p^n$. By Theorem 1.7 and Corollary 1.26, the field $\mathbb{F}_q$ can be written in the form $\mathbb{F}[x]/(f)$, where $f \in \mathbb{F}[x]$ is an irreducible polynomial with degree n. In practical applications of field theory, for example in computer science, this is the most common method of constructing a finite field. Using, for instance, the polynomial $f(x) = x^3 + x + 1$ in Example 1.2., we may construct the field $\mathbb{F}_8$. The following theorem shows that we have a good chance of obtaining an irreducible polynomial by a random selection.

Theorem 1.31 *Let $f(x) \in \mathbb{F}_q[x]$ be a uniformly distributed random polynomial with degree $k > 1$ and leading coefficient 1. (Being uniformly distributed means that the probability of choosing f is $1/q^k$.) Then f is irreducible over $\mathbb{F}_q$ with probability at least $1/k - 1/q^{k/2}$.*

Proof. First we estimate the number of elements $\alpha \in \mathbb{F}_{q^k}$ for which $\mathbb{F}_q(\alpha) = \mathbb{F}_{q^k}$. We claim that the number of such elements is at least

$$|\mathbb{F}_{q^k}| - \sum_{r|k} |\mathbb{F}_{q^{k/r}}| \,,$$

where the summation runs for the distinct prime divisors r of k. Indeed, if α does not generate, over $\mathbb{F}_q$, the field $\mathbb{F}_{q^k}$, then it is contained in a maximal subfield of $\mathbb{F}_{q^k}$, and these maximal subfields are, by Theorem 1.27, exactly the fields of the form $\mathbb{F}_{q^{k/r}}$. The number of distinct prime divisors of k are at most $\lg k$, and so the number of such elements α is at least $q^k - (\lg k)q^{k/2}$. The minimal polynomials with leading coefficients 1 over $\mathbb{F}_q$ of such elements α have degree k and they are irreducible. Such a polynomial is a minimal polynomial of exactly k elements α (Theorem 1.28). Hence the number of distinct irreducible polynomials with degree k and leading coefficient 1 in $\mathbb{F}_q[x]$ is at least

$$\frac{q^k}{k} - \frac{(\lg k)q^{k/2}}{k} \geq \frac{q^k}{k} - q^{k/2} \,,$$

from which the claim follows. ■

If, having $\mathbb{F}_q$, we would like to construct one of its extensions $\mathbb{F}_{q^k}$, then it is worth selecting a ***random*** polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_{k-1}x^{k-1} + x^k \in \mathbb{F}_q[x].$$

In other words, we select uniformly distributed random coefficients $a_0, \ldots, a_{k-1} \in \mathbb{F}_q$ independently. The polynomial so obtained is irreducible with a high probability (in fact, with probability at least $1/k - \epsilon$ if q^k is large). Further, in this case, we also have $\mathbb{F}_q[x]/(f) \cong \mathbb{F}_{q^k}$. We expect that we will have to select about k polynomials before we find an irreducible one.

We have seen in Theorem 1.2 that field extensions can be obtained using irreducible polynomials. It is often useful if these polynomials have some further nice properties. The following lemma claims the existence of such polynomials.

Lemma 1.32 *Let r be a prime. In a finite field $\mathbb{F}_q$ there exists an element which is not an r-th power if and only if $q \equiv 1 \pmod{r}$. If $b \in \mathbb{F}_q$ is such an element, then the polynomial $x^r - b$ is irreducible in $\mathbb{F}_q[x]$, and so $\mathbb{F}_q[x]/(x^r - b)$ is a field with q^r elements.*

Proof. Suppose first that $r \nmid q-1$ and let s be a positive integer such that $sr \equiv 1 \pmod{q-1}$. If $b \in \mathbb{F}_q$ such that $b \neq 0$, then $(b^s)^r = b^{sr} = bb^{sr-1} = b$, while if $b = 0$, then $b = 0^r$. Hence,

in this case, each of the elements of $\mathbb{F}_q$ is an r-th power.

Next we assume that $r \mid q-1$, and we let a be a primitive element in $\mathbb{F}_q$. Then, in $\mathbb{F}_q$, the r-th powers are exactly the following $1+(q-1)/r$ elements: $0, (a^r)^0, (a^r)^1, \ldots, (a^r)^{(q-1)/r-1}$. Suppose now that $r^s \mid q-1$, but $r^{s+1} \nmid q-1$. Then the order of an element $b \in \mathbb{F}_q \setminus \{0\}$ is divisible by r^s if and only if b is not an r-th power. Let b be such an element, and let $g(x) \in \mathbb{F}_q[x]$ be an irreducible factor of the polynomial $x^r - b$. Suppose that the degree of $g(x)$ is d; clearly, $d \le r$. Then $\mathbb{K} = \mathbb{F}_q[x]/(g(x))$ is a field with q^d elements and, in $\mathbb{K}$, the equation $[x]^r = b$ holds. Therefore the order of $[x]$ is divisible by r^{s+1}. Consequently, $r^{s+1} \mid q^d - 1$. As $q-1$ is not divisible by r^{s+1}, we have $r \mid (q^d-1)/(q-1) = 1+q+\cdots+q^{d-1}$. In other words $1+q+\ldots+q^{d-1} \equiv 0 \pmod{r}$. On the other hand, as $q \equiv 1 \pmod{r}$, we find $1+q+\cdots+q^{d-1} \equiv d \pmod{r}$, and hence $d \equiv 0 \pmod{r}$, which, since $0 < d \le r$, can only happen if $d = r$. ∎

In certain cases, we can use the previous lemma to boost the probability of finding an irreducible polynomial.

Proposition 1.33 *Let r be a prime such that $r \mid q-1$. Then, for a random element $b \in \mathbb{F}_q^*$, the polynomial $x^r - b$ is irreducible in $\mathbb{F}_q[x]$ with probability at least $1 - 1/r$.*

Proof. Under the conditions, the r-th powers in $\mathbb{F}_q^*$ constitute the cyclic subgroup with order $(q-1)/r$. Thus a random element $b \in \mathbb{F}_q^*$ is an r-th power with probability $1/r$, and hence the assertion follows from Lemma 1.32. ∎

Remark. Assume that $r \mid (q-1)$, and, if $r = 2$, then assume also that $4 \mid (q-1)$. In this case there is an element b in $\mathbb{F}_q$ that is not an r-th power. We claim that that the residue class $[x]$ is not an r-th power in $\mathbb{F}_q[x]/(x^r - b) \cong \mathbb{F}_q^r$. Indeed, by the argument in the proof of Lemma 1.32, it suffices to show that $r^2 \nmid (q^r-1)/(q-1)$. By our assumptions, this is clear if $r = 2$. Now assume that $r > 2$, and write $q \equiv 1 + rt \pmod{r^2}$. Then, for all integers $i \ge 0$, we have $q^i \equiv 1 + irt \pmod{r^2}$, and so, by the assumptions,

$$\frac{q^r-1}{q-1} = 1 + q + \cdots + q^{r-1} \equiv r + \frac{r(r-1)}{2} rt \equiv r \pmod{r^2}\,.$$

Exercises

1.2-1 Show that the polynomial $x^{q+1} - 1$ can be factored as a product of linear factors over the field $\mathbb{F}_{q^2}$.

1.2-2 Show that the polynomial $f(x) = x^4 + x + 1$ is irreducible over $\mathbb{F}_2$, that is, $\mathbb{F}_2[x]/(f) \cong \mathbb{F}_{16}$. What is the order of the element $[x]_f$ in the residue class ring? Is it true that the element $[x]_f$ is primitive in $\mathbb{F}_{16}$?

1.2-3 Determine the irreducible factors of $x^{31} - 1$ over the field $\mathbb{F}_2$.

1.2-4 Determine the subfields of $\mathbb{F}_{3^6}$.

1.2-5 Let a and b be positive integers. Show that there exists a finite field $\mathbb{K}$ containing $\mathbb{F}_q$ such that $\mathbb{F}_{q^a} \subseteq \mathbb{K}$ and $\mathbb{F}_{q^b} \subseteq \mathbb{K}$. What can we say about the number of elements in $\mathbb{K}$?

1.2-6 Show that the number of irreducible polynomials with degree k and leading coefficient 1 over $\mathbb{F}_q$ is at most q^k/k.

1.2-7 *(a)* Let $\mathbb{F}$ be a field, let V be an n-dimensional vector space over $\mathbb{F}$, and let $A : V \to V$ be a linear transformation whose minimal polynomial coincides with its characteristic polynomial. Show that there exists a vector $v \in V$ such that the images $v, Av, \ldots, A^{n-1}v$ are linearly independent.

(b) A set $S = \{\alpha, \alpha^q, \ldots, \alpha^{q^{d-1}}\}$ is said to be a ***normal basis*** of $\mathbb{F}_{q^d}$ over $\mathbb{F}_q$, if $\alpha \in \mathbb{F}_{q^d}$ and S is a linearly independent set over $\mathbb{F}_q$. Show that $\mathbb{F}_{q^d}$ has a normal basis over $\mathbb{F}_q$. *Hint:* Show that a minimal polynomial of the $\mathbb{F}_q$-linear map $\Phi : \mathbb{F}_{q^d} \to \mathbb{F}_{q^d}$ is $x^d - 1$, and use part (a).

1.3. Factoring polynomials over finite fields

One of the problems that we often have to solve when performing symbolic computation is the ***factorisation*** problem. Factoring an algebraic expression means writing it as a product of simpler expressions. Experience shows that this can be very helpful in the solution of a large variety of algebraic problems. In this section, we consider a class of factorisation algorithms that can be used to factor polynomials in one variable over finite fields.

The input of the ***polynomial factorisation problem*** is a polynomial $f(x) \in \mathbb{F}_q[x]$. Our aim is to compute a factorisation

$$f = f_1^{e_1} f_2^{e_2} \cdots f_s^{e_s} \tag{1.3}$$

of f where the polynomials $f_1, \ldots, f_s$ are pairwise relatively prime and irreducible over $\mathbb{F}_q$, and the exponents e_i are positive integers. By Theorem 1.4, f determines the polynomials f_i and the exponents e_i essentially uniquely.

1.3. Example. Let $p = 23$ and let

$$f(x) = x^6 - 3x^5 + 8x^4 - 11x^3 + 8x^2 - 3x + 1 \,.$$

Then it is easy to compute modulo 23 that

$$f(x) = (x^2 - x + 10)(x^2 + 5x + 1)(x^2 - 7x + 7) \,.$$

None of the factors $x^2 - x + 10$, $x^2 + 5x + 1$, $x^2 - 7x + 7$ has a root in $\mathbb{F}_{23}$, and so they are necessarily irreducible in $\mathbb{F}_{23}[x]$.

The factorisation algorithms are important computational tools, and so they are implemented in most of the computer algebra systems (Mathematica, Maple, etc). These algorithms are often used in the area of error-correcting codes and in cryptography.

Our aim in this section is to present some of the basic ideas and building blocks that can be used to factor polynomials over finite fields. We will place an emphasis on the existence of polynomial time algorithms. The discussion of the currently best known methods is, however, outside the scope of this book.

1.3.1. Square-free factorisation

The factorisation problem in the previous section can efficiently be reduced to the special case when the polynomial f to be factored is square-free; that is, in (1.3), $e_i = 1$ for all i. The basis of this reduction is Lemma 1.13 and the following simple result. Recall that the derivative of a polynomial $f(x)$ is denoted by $f'(x)$.

Lemma 1.34 *Let $f(x) \in \mathbb{F}_q[x]$ be a polynomial. If $f'(x) = 0$, then there exists a polynomial $g(x) \in \mathbb{F}_q[x]$ such that $f(x) = g(x)^p$.*

Proof. Suppose that $f(x) = \sum_{i=0}^{n} a_i x^i$. Then $f'(x) = \sum_{i=1}^{n} a_i i x^{i-1}$. If the coefficient $a_i i$ is zero in $\mathbb{F}_q$ then either $a_i = 0$ or $p \mid i$. Hence, if $f'(x) = 0$ then $f(x)$ can be written as $f(x) = \sum_{j=0}^{k} b_j x^{pj}$. Let $q = p^d$; then choosing $c_j = b_j^{p^{d-1}}$, we have $c_j^p = b_j^{p^d} = b_j$, and so $f(x) = (\sum_{j=0}^{k} c_j x^j)^p$. ■

If $f'(x) = 0$, then, using the previous lemma, a factorisation of $f(x)$ into square-free factors can be obtained from that of the polynomial $g(x)$, which has smaller degree. On the other hand, if $f'(x) \neq 0$, then, by Lemma 1.13, the polynomial $f(x)/\gcd(f(x), f'(x))$ is already square-free and we only have to factor $\gcd(f(x), f'(x))$ into square-free factors. The division of polynomials and computing the greatest common divisor can be performed in polynomial time, by Theorem 1.12. In order to compute the polynomial $g(x)$, we need the solutions, in $\mathbb{F}_q$, of equations of the form $y^p = a$ with $a \in \mathbb{F}_q$. If $q = p^s$, then $y = a^{p^{s-1}}$ is a solution of such an equation, which, using ***fast exponentiation*** (repeated squaring, see 33.6.1), can be obtained in polynomial time.

Reference to NA!

One of the two reduction steps can always be performed if f is divisible by a square of a polynomial with positive degree.

Usually a polynomial can be written as a product of square-free factors in many different ways. For the sake of uniqueness, we define the ***square-free factorisation*** of a polynomial $f \in \mathbb{F}[x]$ as the factorisation

$$f = f_1^{e_1} \cdots f_s^{e_s} ,$$

where $e_1 < \cdots < e_s$ are integers, and the polynomials f_i are relatively prime and square-free. Hence we collect together the irreducible factors of f with the same multiplicity. The following algorithm computes a square-free factorisation of f. Besides the observations we made in this section, we also use Lemma 1.14. This lemma, combined with Lemma 1.13, guarantees that the product of the irreducible factors with multiplicity one of a polynomial f over a finite field is $f/\gcd(f, f')$.

SQUARE-FREE-FACTORISATION(f)

```
1  g ← f
2  S ← ∅
3  m ← 1
4  i ← 1
5  while deg g ≠ 0
6        do if g' = 0
7              then g ← ∛g
8                   i ← i · p
9              else h ← g/gcd(g, g')
10                  g ← g/h
11                  if deg h ≠ 0
12                     then S ← S ∪ (h, m)
13                  m ← m + i
14 return S
```

The degree of the polynomial g decreases after each execution of the main loop, and the subroutines used in this algorithm run in polynomial time. Thus the method above can

be performed in polynomial time.

1.3.2. Distinct degree factorisation

Suppose that f is a square-free polynomial. Now we factor f as

$$f(x) = h_1(x)h_2(x)\cdots h_t(x)\ , \tag{1.4}$$

where, for $i = 1, \ldots, t$, the polynomial $h_i(x) \in \mathbb{F}_q[x]$ is a product of irreducible polynomials with degree i. Though this step is not actually necessary for the solution of the factorisation problem, it is worth considering, as several of the known methods can efficiently exploit the structure of the polynomials h_i. The following fact serves as the starting point of the distinct degree factorisation.

Theorem 1.35 *The polynomial $x^{q^d} - x$ is the product of all the irreducible polynomials $f \in \mathbb{F}_q[x]$, each of which is taken with multiplicity* 1*, that have leading coefficient* 1 *and whose degree divides d.*

Proof. As $(x^{q^d} - x)' = -1$, all the irreducible factors of this polynomial occur with multiplicity one. If $f \in \mathbb{F}_q[x]$ is irreducible and divides $x^{q^d} - x$, then, by Theorem 1.24, the degree of f divides d.

Conversely, let $f \in \mathbb{F}_q[x]$ be an irreducible polynomial with degree k such that $k \mid d$. Then, by Theorem 1.27, f has a root in $\mathbb{F}_{q^d}$, which implies $f \mid x^{q^d} - x$. ∎

The theorem offers an efficient method for computing the polynomials $h_i(x)$. First we separate h_1 from f, and then, step by step, we separate the product of the factors with higher degrees.

DISTINCT-DEGREE-FACTORISATION(f)

```
1  F ← f
3  for i ← 1 to deg f
4      do h_i ← gcd(F, x^{q^i} − x)
7         F ← F/h_i
8  return h_1, ..., h_{deg f}
```

If, in this algorithm, the polynomial $F(x)$ is constant, then we may stop, as the further steps will not give new factors. As the polynomial $x^{q^i} - x$ may have large degree, computing $\gcd(F(x), x^{q^i} - x)$ must be performed with particular care. The important idea here is that the residue $x^{q^i} \pmod{F(x)}$ can be computed using fast exponentiation.

The algorithm outlined above is suitable for testing whether a polynomial is irreducible, which is one of the important problems that we encounter when constructing finite fields. The algorithm presented here for distinct degree factorisation can solve this problem efficiently. For, it is obvious that a polynomial f with degree k is irreducible, if, in the factorisation (1.4), we have $h_k(x) = f(x)$.

The following algorithm for testing whether a polynomial is irreducible is somewhat more efficient than the one sketched in the previous paragraph and handles correctly also the inputs that are not square-free.

IRREDUCIBILITY-TEST(f)

```
1  n ← deg f
2  if x^{p^n} ≢ x (mod f)
3     then return "no"
4  for the prime divisors r of n
5     do if x^{p^{n/r}} ≡ x (mod f)
6           then return "no"
7  return "yes"
```

In lines 2 and 5, we check whether n is the smallest among the positive integers k for which f divides $x^{q^k} - x$. By Theorem 1.35, this is equivalent to the irreducibility of f. If f survives the test in line 2, then, by Theorem 1.35, we know that f is square-free and k must divide n. Using at most $\lg n + 1$ fast exponentiations modulo f, we can thus decide if f is irreducible.

Theorem 1.36 *If the field $\mathbb{F}_q$ is given and $k > 1$ is an integer, then the field $\mathbb{F}_{q^k}$ can be constructed using a randomised Las Vegas algorithm which runs in time polynomial in $\lg q$ and k.*

Proof. The algorithm is the following.

FINITE-FIELD-CONSTRUCTION(q^k)

```
1  for i ← 0 to k − 1
2     do a_i ← a random element (uniformly distributed) of F_q
3  f ← x^k + Σ_{i=0}^{k−1} a_i x^i
4  if IRREDUCIBILITY-TEST(f) = "yes"
5     then return F_q[x]/(f)
6     else return "fail"
```

In lines 1–3, we choose a uniformly distributed random polynomial with leading coefficient 1 and degree k. Then, in line 4, we efficiently check if $f(x)$ is irreducible. By Theorem 1.31, the polynomial f is irreducible with a reasonably high probability. ■

1.3.3. The Cantor-Zassenhaus algorithm

In this section we consider the special case of the factorisation problem in which q is odd and the polynomial $f(x) \in \mathbb{F}_q[x]$ is of the form

$$f = f_1 f_2 \cdots f_s, \tag{1.5}$$

where the f_i are pairwise relatively prime irreducible polynomials in $\mathbb{F}_q[x]$ with the same degree d, and we also assume that $s \geq 2$. Our motivation for investigating this special case is that a square-free distinct degree factorisation reduces the general factorisation problem to such a simpler problem. If q is even, then Berlekamp's method, presented in Section 1.3.4, gives a deterministic polynomial time solution. There is a variation of the method discussed in the present section that works also for even q; see Exercise *1-2.*.

The second dot after Exercise 18-2 comes from a macro!

Lemma 1.37 *Suppose that q is odd. Then there are $(q^2-1)/2$ pairs $(c_1, c_2) \in \mathbb{F}_q \times \mathbb{F}_q$ such that exactly one of $c_1^{(q-1)/2}$ and $c_2^{(q-1)/2}$ is equal to 1.*

Proof. Suppose that a is a primitive element in $\mathbb{F}_q$; that is, $a^{q-1} = 1$, but $a^k \neq 1$ for $0 < k < q-1$. Then $\mathbb{F}_q \setminus \{0\} = \{a^s | s = 0, \ldots, q-2\}$, and further, as $\left(a^{(q-1)/2}\right)^2 = 1$, but $a^{(q-1)/2} \neq 1$, we obtain that $a^{(q-1)/2} = -1$. Therefore $a^{s(q-1)/2} = (-1)^s$, and so half of the element $c \in \mathbb{F}_q \setminus \{0\}$ give $c^{(q-1)/2} = 1$, while the other half give $c^{(q-1)/2} = -1$. If $c = 0$ then clearly $c^{(q-1)/2} = 0$. Thus there are $((q-1)/2)((q+1)/2)$ pairs (c_1, c_2) such that $c_1^{(q-1)/2} = 1$, but $c_2^{(q-1)/2} \neq 1$, and, obviously, we have the same number of pairs for which the converse is valid. Thus the number of pairs that satisfy the condition is $(q-1)(q+1)/2 = (q^2-1)/2$. ∎

Theorem 1.38 *Suppose that q is odd and the polynomial $f(x) \in \mathbb{F}_q[x]$ is of the form (1.5) and has degree n. Choose a uniformly distributed random polynomial $u(x) \in \mathbb{F}_q[x]$ with degree less than n. (That is, choose pairwise independent, uniformly distributed scalars $u_0, \ldots, u_{n-1}$, and consider the polynomial $u(x) = \sum_{i=0}^{n-1} u_i x^i$.) Then, with probability at least $(q^{2d}-1)/(2q^{2d}) \geq 4/9$, the greatest common divisor*

$$\gcd(u(x)^{\frac{q^d-1}{2}} - 1, f(x))$$

is a proper divisor of $f(x)$.

Proof. The element $u(x) \pmod{f_i(x)}$ corresponds to an element of the residue class field $\mathbb{F}[x]/(f_i(x)) \cong \mathbb{F}_{q^d}$. By the Chinese remainder theorem (Theorem 1.15), choosing the polynomial $u(x)$ uniformly implies that the residues of $u(x)$ modulo the factors $f_i(x)$ are independent and uniformly distributed random polynomials. By Lemma 1.37, the probability that exactly one of the residues of the polynomial $u(x)^{(q^d-1)/2} - 1$ modulo $f_1(x)$ and $f_2(x)$ is zero is precisely $(q^{2d}-1)/(2q^{2d})$. In this case the greatest common divisor in the theorem is indeed a divisor of f. For, if $u(x)^{(q^d-1)/2} - 1 \equiv 0 \pmod{f_1(x)}$, but this congruence is not valid modulo $f_2(x)$, then the polynomial $u(x)^{(q^d-1)/2} - 1$ is divisible by the factor $f_1(x)$, but not divisible by $f_2(x)$, and so its greatest common divisor with $f(x)$ is a proper divisor of $f(x)$. The function

$$\frac{q^{2d}-1}{2q^{2d}} = \frac{1}{2} - \frac{1}{2q^{2d}}$$

is strictly increasing in q^d, and it takes its smallest possible value if q^d is the smallest odd prime-power, namely 3. The minimum is, thus, $1/2 - 1/18 = 4/9$. ∎

The previous theorem suggests the following randomised Las Vegas polynomial time algorithm for factoring a polynomial of the form (1.5) to a product of two factors.

CANTOR-ZASSENHAUS-ODD(f, d)

```
1  n ← deg f
2  for i ← 0 to n − 1
3      do u_i ← a random element (uniformly distributed) of F_q
4  u ← Σ_{i=0}^{n−1} u_i x^i
5  g ← gcd(u^{(q^d−1)/2} − 1, f)
6  if 0 < deg g < deg f
7     then return(g, f/g)
8     else return "fail"
```

If one of the polynomials in the output is not irreducible, then, as it is of the form (1.5), it can be fed, as input, back into the algorithm. This way we obtain a polynomial time randomised algorithm for factoring f.

In the computation of the greatest common divisor, the residue $u(x)^{(q^d-1)/2} \pmod{f(x)}$ should be computed using fast exponentiation.

Now we can conclude that the general factorisation problem (1.3) over a field with odd order can be solved using a randomised polynomial time algorithm.

1.3.4. Berlekamp's algorithm

Here we will describe an algorithm that reduces the problem of factoring polynomials to the problem of searching through the underlying field or its prime field. We assume that

$$f(x) = f_1^{e_1}(x) \cdots f_s^{e_s}(x) \ ,$$

where the $f_i(x)$ are pairwise non-associate, irreducible polynomials in $\mathbb{F}_q[x]$, and also that $\deg f(x) = n$. The Chinese remainder theorem (Theorem 1.15) gives an isomorphism between the rings $\mathbb{F}_q[x]/(f)$ and

$$\mathbb{F}_q[x]/(f_1^{e_1}) \oplus \cdots \oplus \mathbb{F}_q[x]/(f_s^{e_s}).$$

The isomorphism is given by the following map:

$$[u(x)]_f \leftrightarrow ([u(x)]_{f_1^{e_1}}, \ldots, [u(x)]_{f_s^{e_s}}) \ ,$$

where $u(x) \in \mathbb{F}_q[x]$.

The most important technical tools in Berlekamp's algorithm are the p-th and q-th power maps in the residue class ring $\mathbb{F}_q[x]/(f(x))$. Taking p-th and q-th powers on both sides of the isomorphism above given by the Chinese remainder theorem, we obtain the following maps:

$$[u(x)]^p \quad \leftrightarrow \quad ([u(x)^p]_{f_1^{e_1}}, \ldots, [u(x)^p]_{f_s^{e_s}}) \ , \tag{1.6}$$

$$[u(x)]^q \quad \leftrightarrow \quad ([u(x)^q]_{f_1^{e_1}}, \ldots, [u(x)^q]_{f_s^{e_s}}) \ . \tag{1.7}$$

The ***Berlekamp subalgebra*** B_f of the polynomial $f = f(x)$ is the subring of the residue class ring $\mathbb{F}_q[x]/(f)$ consisting of the fixed points of the q-th power map. Further, the ***absolute Berlekamp subalgebra*** A_f of f consists of the fixed points of the p-th power map. In symbols,

$$B_f = \{[u(x)]_f \in \mathbb{F}_q[x]/(f) : \ [u(x)^q]_f = [u(x)]_f\} \ ,$$

$$A_f = \{[u(x)]_f \in \mathbb{F}_q[x]/(f) : \ [u(x)^p]_f = [u(x)]_f\} \ .$$

It is easy to see that $A_f \subseteq B_f$. The term subalgebra is used here, because both types of Berlekamp subalgebras are subrings in the residue class ring $\mathbb{F}_q[x]/(f(x))$ (that is they are closed under addition and multiplication modulo $f(x)$), and, in addition, B_f is also linear

subspace over $\mathbb{F}_q$, that is, it is closed under multiplication by the elements of $\mathbb{F}_q$. The absolute Berlekamp subalgebra A_f is only closed under multiplication by the elements of the prime field $\mathbb{F}_p$.

The Berlekamp subalgebra B_f is a subspace, as the map $u \mapsto u^q - u \pmod{f(x)}$ is an $\mathbb{F}_q$-linear map of $\mathbb{F}_q[x]/g(x)$ into itself, by Lemma 1.23 and Theorem 1.19. Hence a basis of B_f can be computed as a solution of a homogeneous system of linear equations over $\mathbb{F}_q$, as follows.

For all $i \in \{0, \ldots, n-1\}$, compute the polynomial $h_i(x)$ with degree at most $n-1$ that satisfies $x^{iq} - x^i \equiv h_i(x) \pmod{f(x)}$. For each i, such a polynomial h_i can be determined by fast exponentiation using $O(\lg q)$ multiplications of polynomials and divisions with remainder. Set $h_i(x) = \sum_{j=0}^{n} h_{ij} x^j$. The class $[u]_f$ of a polynomial $u(x) = \sum_{i=0}^{n-1} u_i x^i$ with degree less than n lies in the Berlekamp subalgebra if and only if

$$\sum_{i=0}^{n-1} u_i h_i(x) = 0 \,,$$

which, considering the coefficient of x^j for $j = 0, \ldots, n-1$, leads to the following system of n homogeneous linear equations in n variables:

$$\sum_{i=0}^{n-1} h_{ij} u_i = 0, \quad (j = 0, \ldots, n-1) \,.$$

Similarly, computing a basis of the absolute Berlekamp subalgebra over $\mathbb{F}_p$ can be carried out by solving a system of nd homogeneous linear equations in nd variables over the prime field $\mathbb{F}_p$, as follows. We represent the elements of $\mathbb{F}_q$ in the usual way, namely using polynomials with degree less than d in $\mathbb{F}_p[y]$. We perform the operations modulo $g(y)$, where $g(y) \in \mathbb{F}_p[y]$ is an irreducible polynomial with degree d over the prime field $\mathbb{F}_p$. Then the polynomial $u[x] \in \mathbb{F}_q[x]$ of degree less than n can be written in the form

$$\sum_{i=0}^{n-1} \sum_{j=0}^{d-1} u_{ij} y^j x^i,$$

where $u_{ij} \in \mathbb{F}_p$. Let, for all $i \in \{0, \ldots, n-1\}$ and for all $j \in \{0, \ldots, d-1\}$, $h_{ij}(x) \in \mathbb{F}_q[x]$ be the unique polynomial with degree at most $(n-1)$ for which $h_{ij}(x) \equiv (y^j x^i)^p - y^j x^i \pmod{f(x)}$. The polynomial $h_{ij}(x)$ is of the form $\sum_{k=0}^{n-1} \sum_{l=0}^{d-1} h_{ij}^{kl} y^l x^k$. The criterion for being a member of the absolute Berlekamp subalgebra of $[u]$ with $u[x] = \sum_{i=0}^{n-1} \sum_{j=0}^{d-1} u_{ij} y^j x^i$ is

$$\sum_{i=0}^{n-1} \sum_{j=0}^{d-1} u_{ij} h_{ij}(x) = 0 \,,$$

which, considering the coefficients of the monomials $y^l x^k$, is equivalent to the following system of equations:

$$\sum_{i=0}^{n-1} \sum_{j=0}^{d-1} h_{ij}^{kl} u_{ij} = 0 \quad (k = 0, \ldots, n-1,\ l = 0, \ldots, d-1) \,.$$

This is indeed a homogeneous system of linear equations in the variables u_{ij}. Systems of

linear equations over fields can be solved in polynomial time (see Section 31.4), the operations in the ring $\mathbb{F}_q[x]/(f(x))$ can be performed in polynomial time, and the fast exponentiation also runs in polynomial time. Thus the following theorem is valid.

Reference to NA!

Theorem 1.39 *Let $f \in \mathbb{F}_q[x]$. Then it is possible to compute the Berlekamp subalgebras $B_f \leq \mathbb{F}_q[x]/(f(x))$ and $A_f \leq \mathbb{F}_q[x]/(f(x))$, in the sense that an $\mathbb{F}_q$-basis of B_f and $\mathbb{F}_p$-basis of A_f can be obtained, using polynomial time deterministic algorithms.*

By (1.6) and (1.7),

$$B_f = \{[u(x)]_f \in \mathbb{F}_q[x]/(f) : \ [u^q(x)]_{f_i^{e_i}} = [u(x)]_{f_i^{e_i}} \ (i = 1, \ldots, s)\} \tag{1.8}$$

and

$$A_f = \{[u(x)]_f \in \mathbb{F}_q[x]/(f) : \ [u^p(x)]_{f_i^{e_i}} = [u(x)]_{f_i^{e_i}} \ (i = 1, \ldots, s)\} \ . \tag{1.9}$$

The following theorem shows that the elements of the Berlekamp subalgebra can be characterised by their Chinese remainders.

Theorem 1.40

$$B_f = \{[u(x)]_f \in \mathbb{F}_q[x]/(f) : \ \exists c_i \in \mathbb{F}_q \text{ such that } [u(x)]_{f_i^{e_i}} = [c_i]_{f_i^{e_i}} \ (i = 1, \ldots, s)\}$$

and

$$A_f = \{[u(x)]_f \in \mathbb{F}_q[x]/(f) : \ \exists c_i \in \mathbb{F}_p \text{ such that } [u(x)]_{f_i^{e_i}} = [c_i]_{f_i^{e_i}} \ (i = 1, \ldots, s)\} \ .$$

Proof. Using the Chinese remainder theorem, and equations (1.8), (1.9), we are only required to prove that

$$u^q(x) \equiv u(x) \pmod{g^e(x)} \Longleftrightarrow \exists c \in \mathbb{F}_q \text{ such that } u(x) \equiv c \pmod{g^e(x)} \ ,$$

and

$$u^p(x) \equiv u(x) \pmod{g^e(x)} \Longleftrightarrow \exists c \in \mathbb{F}_p \text{ such that } u(x) \equiv c \pmod{g^e(x)}$$

where $g(x) \in \mathbb{F}_q[x]$ is an irreducible polynomial, $u(x) \in \mathbb{F}_q[x]$ is an arbitrary polynomial and e is a positive integer. In both of the cases, the direction $\Leftarrow$ is a simple consequence of Theorem 1.19. As $\mathbb{F}_p = \{a \in \mathbb{F}_q \mid a^p = a\}$, the implication $\Rightarrow$ concerning the absolute Berlekamp subalgebra follows from that concerning the Berlekamp subalgebra, and so it suffices to consider the latter.

The residue class ring $\mathbb{F}_q[x]/(g(x))$ is a field, and so the polynomial $x^q - x$ has at most q roots in $\mathbb{F}_q[x]/(g(x))$. However, we already obtain q distinct roots from Theorem 1.19, namely the elements of $\mathbb{F}_q$ (the constant polynomials modulo $g(x)$). Thus

$$u^q(x) \equiv u(x) \pmod{g(x)} \Longleftrightarrow \exists c \in \mathbb{F}_q \text{ such that } u(x) \equiv c \pmod{g(x)} \ .$$

Hence, if $u^q(x) \equiv u(x) \pmod{g^e(x)}$, then $u(x)$ is of the form $u(x) = c + h(x)g(x)$ where $h(x) \in \mathbb{F}_q[x]$. Let N be an arbitrary positive integer. Then

$$u(x) \equiv u^q(x) \equiv u^{q^N}(x) \equiv (c + h(x)g(x))^{q^N} \equiv c + h(x)^{q^N} g(x)^{q^N} \equiv c \pmod{g^{q^N}(x)} \ .$$

If we choose N large enough so that $q^N \geq e$ holds, then, by the congruence above, $u(x) \equiv c \pmod{g^e(x)}$ also holds. ■

An element $[u(x)]_f$ of B_f or A_f is said to be ***non-trivial*** if there is no element $c \in \mathbb{F}_q$ such that $u(x) \equiv c \pmod{f(x)}$. By the previous theorem and the Chinese remainder theorem, this holds if and only if there are $i,\ j$ such that $c_i \neq c_j$. Clearly a necessary condition is that $s > 1$, that is, $f(x)$ must have at least two irreducible factors.

Lemma 1.41 *Let $[u(x)]_f$ be a non-trivial element of the Berlekamp subalgebra B_f. Then there is an element $c \in \mathbb{F}_q$ such that the polynomial* $\gcd(u(x) - c, f(x))$ *is a proper divisor of $f(x)$. If $[u(x)]_f \in A_f$, then there exists such an element c in the prime field $\mathbb{F}_p$.*

Proof. Let i and j be integers such that $c_i \neq c_j \in \mathbb{F}_q$, $u(x) \equiv c_i \pmod{f_i^{e_i}(x)}$, and $u(x) \equiv c_j \pmod{f_j^{e_j}(x)}$. Then, choosing $c = c_i$, the polynomial $u(x) - c$ is divisible by $f_i^{e_i}(x)$, but not divisible by $f_j^{e_j}(x)$. If, in addition, $u(x) \in A_f$, then also $c = c_i \in \mathbb{F}_p$. ■

Assume that we have a basis of A_f at hand. At most one of the basis elements can be trivial, as a trivial element is a scalar multiple of 1. If $f(x)$ is not a power of an irreducible polynomial, then there will surely be a non-trivial basis element $[u(x)]_f$, and so, using the idea in the previous lemma, $f(x)$ can be factored two factors.

Theorem 1.42 *A polynomial $f(x) \in \mathbb{F}_q[x]$ can be factored with a deterministic algorithm whose running time is polynomial in p,* $\deg f$, *and* $\lg q$.

Proof. It suffices to show that f can be factored to *two* factors within the given time bound. The method can then be repeated.

BERLEKAMP-DETERMINISTIC(f)

```
1  S ← a basis of A_f
2  if |S| > 1
3     then u ← a non-trivial element of S
4     for c ∈ F_p
5         do g ← gcd(u − c, f)
6            if 0 < deg g < deg f
7               then return (g, f/g)
8     else return "a power of an irreducible"
```

In the first stage, in line 1, we determine a basis of the absolute Berlekamp subalgebra. The cost of this is polynomial in $\deg f$ and $\lg q$. In the second stage (lines 2–8), after taking a non-trivial basis element $[u(x)]_f$, we compute the greatest common divisors $\gcd(u(x) - c, f(x))$ for all $c \in \mathbb{F}_p$. The cost of this is polynomial in p and $\deg f$.

If there is no non-trivial basis-element, then A_f is 1-dimensional and f is the e_1-th power of the irreducible polynomial f_1 where f_1 and e_1 can, for instance, be determined using the ideas presented in Section 1.3.1. ■

The time bound in the previous theorem is *not polynomial* in the input size, as it contains p instead of $\lg p$. However, if p is small compared to the other parameters (for instance in coding theory we often have $p = 2$), then the running time of the algorithm will be polynomial in the input size.

Corollary 1.43 *Suppose that p can be bounded by a polynomial function of* $\deg f$ *and*

lg q. *Then the irreducible factorisation of* f *can be obtained in polynomial time.*

The previous two results are due to E. R. Berlekamp. The most important open problem in the area discussed here is the existence of a deterministic polynomial time method for factoring polynomials. The question is mostly of theoretical interest, since the randomised polynomial time methods, such as the a Cantor -Zassenhaus algorithm, are very efficient in practice.

Berlekamp's randomised algorithm

We can obtain a good randomised algorithm using Berlekamp subalgebras. Suppose that q is odd, and, as before, $f \in \mathbb{F}_q[x]$ is the polynomial to be factored.

Let $[u(x)]_f$ be a random element in the Berlekamp subalgebra B_f. An argument, similar to the one in the analysis of the Cantor-Zassenhaus algorithm shows that, provided $f(x)$ has at least two irreducible factors, the greatest common divisor $\gcd(u(x)^{(q-1)/2} - 1, f(x))$ is a proper divisor of $f(x)$ with probability at least 4/9. Now we present a variation of this idea that uses less random bits: instead of choosing a random element from B_f, we only choose a random element from $\mathbb{F}_q$.

Lemma 1.44 *Suppose that* q *is odd and let* a_1 *and* a_2 *be two distinct elements of* $\mathbb{F}_q$. *Then there are at least* $(q-1)/2$ *elements* $b \in \mathbb{F}_q$ *such that exactly one of the elements* $(a_1+b)^{(q-1)/2}$ *and* $(a_2+b)^{(q-1)/2}$ *is* 1.

Proof. Using the argument at the beginning of the proof of Lemma 1.37, one can easily see that there are $(q-1)/2$ elements in the set $\mathbb{F}_q \setminus \{1\}$ whose $(q-1)/2$-th power is -1. It is also quite easy to check, for a given element $c \in \mathbb{F}_q \setminus \{1\}$, that there is a unique $b \neq -a_2$ such that $c = (a_1+b)/(a_2+b)$. Indeed, the required b is the solution of a linear equation.

By the above, there are $(q-1)/2$ elements $b \in \mathbb{F}_q \setminus \{-a_2\}$ such that

$$\left(\frac{a_1+b}{a_2+b}\right)^{(q-1)/2} = -1 \ .$$

For such a b, one of the elements $(a_1+b)^{(q-1)/2}$ and $(a_2+b)^{(q-1)/2}$ is equal to 1 and the other is equal to -1. ∎

Theorem 1.45 *Suppose that* q *is odd and the polynomial* $f(x) \in \mathbb{F}_q[x]$ *has at least two irreducible factors in* $\mathbb{F}_q[x]$. *Let* $u(x)$ *be a non-trivial element in the Berlekamp subalgebra* B_f. *If we choose a uniformly distributed random element* $b \in \mathbb{F}_q$, *then, with probability at least* $(q-1)/(2q) \geq 1/3$, *the greatest common divisor* $\gcd((u(x)+b)^{(q-1)/2} - 1, f(x))$ *is a proper divisor of the polynomial* $f(x)$.

Proof. Let $f(x) = \prod_{i=1}^{s} f_i^{e_i}(x)$, where the factors $f_i(x)$ are pairwise distinct irreducible polynomials. The element $[u(x)]_f$ is a non-trivial element of the Berlekamp subalgebra, and so there are indices $0 < i, j \leq s$ and elements $c_i \neq c_j \in \mathbb{F}_q$ such that $u(x) \equiv c_i \pmod{f_i^{e_i}(x)}$ and $u(x) \equiv c_j \pmod{f_j^{e_j}(x)}$. Using Lemma 1.44 with $a_1 = c_i$ and $a_2 = c_j$, we find, for a random element $b \in \mathbb{F}_q$, that the probability that exactly one of the elements $(c_i+b)^{(q-1)/2} - 1$ and $(c_j+b)^{(q-1)/2} - 1$ is zero is at least $(q-1)/(2q)$. If, for instance, $(c_i+b)^{(q-1)/2} - 1 = 0$, but $(c_j+b)^{(q-1)/2} - 1 \neq 0$, then $(u(x)+b)^{(q-1)/2} - 1 \equiv 0 \pmod{f_i^{e_i}(x)}$ but $(u(x)+b)^{(q-1)/2} - 1 \neq 0$

(mod $f_j^{e_j}(x)$), that is, the polynomial $(u(x)+b)^{(q-1)/2}-1$ is divisible by $f_i^{e_i}(x)$, but not divisible by $f_j^{e_j}(x)$. Thus the greatest common divisor $\gcd(f(x),(u(x)+b)^{(q-1)/2}-1)$ is a proper divisor of f.

The quantity $(q-1)/(2q) = 1/2 - 1/(2q)$ is a strictly increasing function in q, and so it takes its smallest value for the smallest odd prime-power, namely 3. The minimum is 1/3. ■

The previous theorem gives the following algorithm for factoring a polynomial to two factors.

BERLEKAMP-RANDOMISED(f)

```
1  S ← a basis of B_f
2  if |S| > 1
3     then u ← a non-trivial elements of S
4          c ← a random element (uniformly distributed) of F_q
5          g ← gcd((u − c)^{(q−1)/2} − 1, f)
6          if 0 < deg g < deg f
7             then return (g, f/g)
8             else return "fail"
9     else return "a power of an irreducible"
```

Exercises

1.3-1 Let $f(x) \in \mathbb{F}_p[x]$ be an irreducible polynomial, and let α be an element of the field $\mathbb{F}_p[x]/(f(x))$. Give a polynomial time algorithm for computing α^{-1}. *Hint:* Use the result of Exercise 1.1-6.

1.3-2 Let $f(x) = x^7+x^6+x^5+x^4+x^3+x^2+x+1 \in \mathbb{F}_2[x]$. Using the DISTINCT-DEGREE-FACTORISATION algorithm, determine the factorisation (1.4) of f.

1.3-3 Follow the steps of the Cantor-Zassenhaus algorithm to factor the polynomial $x^2+2x+9 \in \mathbb{F}_{11}[x]$.

1.3-4 Let $f(x) = x^2-3x+2 \in \mathbb{F}_5[x]$. Show that $\mathbb{F}_5[x]/(f(x))$ coincides with the absolute Berlekamp subalgebra of f, that is, $A_f = \mathbb{F}_5[x]/(f(x))$.

1.3-5 Let $f(x) = x^3-x^2+x-1 \in \mathbb{F}_7[x]$. Using Berlekamp's algorithm, determine the irreducible factors of f: first find a non-trivial element in the Berlekamp subalgebra A_f, then use it to factor f.

1.4. Lattice reduction

Our aim in the rest of this chapter is to present the Lenstra-Lenstra-Lovász algorithm for factoring polynomials with rational coefficients. First we study a geometric problem, which is interesting also in its own right, namely finding short lattice vectors. Finding a shortest non-zero lattice vector is hard: by a result of Ajtai, if this problem could be solved in polynomial time with a randomised algorithm, then so could all the problems in the complexity class *NP*. For a lattice with dimension n, the lattice reduction method presented in this chapter outputs, in polynomial time, a lattice vector whose length is not greater than $2^{(n-1)/4}$ times the length of a shortest non-zero lattice vector.

1.4.1. Lattices

First, we recall a couple of concepts related to real vector spaces. Let $\mathbb{R}^n$ denote the collection of real vectors of length n. It is routine to check that $\mathbb{R}^n$ is a vector space over the field $\mathbb{R}$. The ***scalar product*** of two vectors $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ in $\mathbb{R}^n$ is defined as the number $(u, v) = u_1v_1 + u_2v_2 + \cdots + u_nv_n$. The quantity $|u| = \sqrt{(u, u)}$ is called the ***length*** of the vector u. The vectors u and v are said to be ***orthogonal*** if $(u, v) = 0$. A basis $b_1, \ldots, b_n$ of the space $\mathbb{R}^n$ is said to be orthonormal, if, for all i, $(b_i, b_i) = 1$ and, for all i and j such that $i \neq j$, we have $(b_i, b_j) = 0$.

The rank and the determinant of a real matrix, and definite matrices are discussed in Section 31.1.

Reference to NA!

Definition 1.46 *A set $L \subseteq \mathbb{R}^n$ is said to be a **lattice**, if L is a subgroup with respect to addition, and L is discrete, in the sense that each bounded region of $\mathbb{R}^n$ contains only finitely many points of L. The **rank** of the lattice L is the dimension of the subspace generated by L. Clearly, the rank of L coincides with the cardinality of a maximal linearly independent subset of L. If L has rank n, then L is said to be a **full** lattice. The elements of L are called **lattice vectors** or **lattice points**.*

Definition 1.47 *Let $b_1, \ldots, b_r$ be linearly independent elements of a lattice $L \subseteq \mathbb{R}^n$. If all the elements of L can be written as linear combinations of the elements $b_1, \ldots, b_r$ with integer coefficients, then the collection $b_1, \ldots, b_r$ is said to be a **basis** of L.*

In this case, as the vectors $b_1, \ldots, b_r$ are linearly independent, all vectors of $\mathbb{R}^n$ can uniquely be written as real linear combinations of $b_1, \ldots, b_r$.

By the following theorem, the lattices are precisely those additive subgroups of $\mathbb{R}^n$ that have bases.

Theorem 1.48 *Let $b_1, \ldots, b_r$ be linearly independent vectors in $\mathbb{R}^n$ and let L be the set of integer linear combinations of $b_1, \ldots, b_r$. Then L is a lattice and the vectors $b_1, \ldots, b_r$ form a basis of L. Conversely, if L is a lattice in $\mathbb{R}^n$, then it has a basis.*

Proof. Obviously, L is a subgroup, that is, it is closed under addition and subtraction. In order to show that it is discrete, let us assume that $n = r$. This assumption means no loss of generality, as the subspace spanned by $b_1, \ldots, b_r$ is isomorphic to $\mathbb{R}^r$. In this case, $\phi : (\alpha_1, \ldots, \alpha_n) \mapsto \alpha_1 b_1 + \ldots + \alpha_n b_n$ is an invertible linear map of $\mathbb{R}^n$ onto itself. Consequently, both ϕ and ϕ^{-1} are continuous. Hence the image of a discrete set under ϕ is also discrete. As $L = \phi(\mathbf{Z}^n)$, it suffices to show that $\mathbf{Z}^n$ is discrete in $\mathbb{R}^n$. This, however, is obvious: if K is a bounded region in $\mathbb{R}^n$, then there is a positive integer ρ, such that the absolute value of each of the coordinates of the elements of K is at most ρ. Thus $\mathbf{Z}^n$ has at most $(2\lfloor\rho\rfloor + 1)^n$ elements in K.

The second assertion is proved by induction on n. If $L = \{0\}$, then we have nothing to prove. Otherwise, by discreteness, there is a shortest non-zero vector, b_1 say, in L. We claim that the vectors of L that lie on the line $\{\lambda b_1 \mid \lambda \in \mathbb{R}\}$ are exactly the integer multiples of b_1. Indeed, suppose that λ is a real number and consider the vector $\lambda b_1 \in L$. As usual, $\{\lambda\}$ denotes the fractional part of λ. Then $0 \neq |\{\lambda\}b_1| < |b_1|$, yet $\{\lambda\}b_1 = \lambda b_1 - [\lambda]b_1$, that is $\{\lambda\}b_1$ is the difference of two vectors of L, and so is itself in L. This, however, contradicts to the fact that b_1 was a shortest non-zero vector in L. Thus our claim holds.

The claim verified in the previous paragraph shows that the theorem is valid when $n = 1$. Let us, hence, assume that $n > 1$. We may write an element of $\mathbb{R}^n$ as the sum of two vectors, one of them is parallel to b_1 and the other one is orthogonal to b_1:

$$v = v^* + \frac{(v, b_1)}{(b_1, b_1)} b_1 \ .$$

Simple computation shows that $(v^*, b_1) = 0$, and the map $v \mapsto v^*$ is linear. Let $L^* = \{v^* | v \in L\}$. We show that L^* is a lattice in the subspace, or hyperplane, $H \cong \mathbb{R}^{n-1}$ formed by the vectors orthogonal to b_1. The map $v \mapsto v^*$ is linear, and so L^* is closed under addition and subtraction. In order to show that it is discrete, let K be a bounded region in H. We are required to show that only finitely many points of L^* are in K. Let $v \in L$ be a vector such that $v^* \in K$. Let λ be the integer that is closest to the number $(v, b_1)/(b_1, b_1)$ and let $v' = v - \lambda b_1$. Clearly, $v' \in L$ and $v'^* = v^*$. Further, we also have that $|(v', b_1)/(b_1, b_1)| = |(v - \lambda b_1, b_1)/(b_1, b_1)| \leq 1/2$, and so the vector v' lies in the bounded region $K \times \{\mu b_1 : \ -1/2 \leq \mu \leq 1/2\}$. However, there are only finitely many vectors $v' \in L$ in this latter region, and so K also has only finitely many lattice vectors $v^* = v'^* \in L^*$.

We have, thus, shown that L^* is a lattice in H, and, by the induction hypothesis, it has a basis. Let $b_2, \ldots, b_r \in L$ be lattice vectors such that the vectors $b_2^*, \ldots, b_r^*$ form a basis of the lattice L^*. Then, for an arbitrary lattice vector $v \in L$, the vector v^* can be written in the form $\sum_{i=2}^r \lambda_i b_i^*$ where the coefficients λ_i are integers. Then $v' = v - \sum_{i=2}^r \lambda_i b_i \in L$ and, as the map $v \mapsto v^*$ is linear, we have $v'^* = 0$. This, however, implies that v' is a lattice vector on the line λb_1, and so $v' = \lambda_1 b_1$ with some integer λ_1. Therefore $v = \sum_{i=1}^r \lambda_i b_i$, that is, v is an integer linear combination of the vectors $b_1, \ldots, b_r$. Thus the vectors $b_1, \ldots, b_r$ form a basis of L. ■

A lattice L is always full in the linear subspace spanned by L. Thus, without loss of generality, we will consider only full lattices, and, in the sequel, by a lattice we will always mean a *full lattice* .

1.4. Example. Two familiar lattices in $\mathbb{R}^2$:
1. The ***square lattice*** is the lattice in $\mathbb{R}^2$ with basis $b_1 = (1, 0),\ b_2 = (0, 1)$.
2. The ***triangular lattice*** is the lattice with basis $b_1 = (1, 0)$, $b_2 = (1/2, (\sqrt{3})/2)$.

The following simple fact will often be used.

Lemma 1.49 *Let L be a lattice in $\mathbb{R}^n$, and let $b_1, \ldots, b_n$ be a basis of L. If we reorder the basis vectors $b_1, \ldots, b_n$, or if we add to a basis vector an integer linear combination of the other basis vectors, then the collection so obtained will also form a basis of L.*

Proof. Straightforward. ■

Let $b_1, \ldots, b_n$ be a basis in L. The Gram matrix of $b_1, \ldots, b_n$ is the matrix $B = (B_{ij})$ with entries $B_{ij} = (b_i, b_j)$. The matrix B is positive definite, since it is of the form $A^T A$ where A is a full-rank matrix (see Theorem 31.6). Consequently, $\det B$ is a positive real number.

Reference to NA!

Lemma 1.50 *Let $b_1, \ldots, b_n$ and $w_1, \ldots, w_n$ be bases of a lattice L and let B and W be the matrices $B_{ij} = (b_i, b_j)$ and $W_{ij} = (w_i, w_j)$. Then the determinants of B and W coincide.*

Proof. For all $i = 1, \ldots, n$, the vector w_i is of the form $w_i = \sum_{j=1}^n \alpha_{ij} b_j$ where the α_{ij} are

integers. Let A be the matrix with entries $A_{ij} = \alpha_{ij}$. Then, as

$$(w_i, w_j) \quad = \quad (\sum_{k=1}^{n} \alpha_{ik} b_k, \sum_{l=1}^{n} \alpha_{jl} b_l) = \sum_{k=1}^{n} \alpha_{ik} \sum_{l=1}^{n} (b_k, b_l) \alpha_{jl},$$

we have $W = ABA^T$, and so $\det W = \det B (\det A)^2$. The number $\det W / \det B = (\det A)^2$ is a non-negative integer, since the entries of A are integers. Swapping the two bases, the same argument shows that $\det B / \det W$ is also a non-negative integer. This can only happen if $\det B = \det W$. ∎

Definition 1.51 (The determinant of a lattice). *The determinant of a lattice L is* $\det L = \sqrt{\det B}$ *where B is the Gram matrix of a basis of L.*

By the previous lemma, $\det L$ is independent of the choice of the basis. The quantity $\det L$ has a geometric meaning, as $\det L$ is the volume of the solid body, the so-called parallelepiped, formed by the vectors $\{\sum_{i=1}^{n} \alpha_i b_i : \ 0 \le \alpha_1, \ldots, \alpha_n \le 1\}$.

Remark 1.52 *Assume that the coordinates of the vectors b_i in an orthonormal basis of $\mathbb{R}^n$ are $\alpha_{i1}, \ldots, \alpha_{in}$ $(i = 1, \ldots, n)$. Then the Gram matrix B of the vectors $b_1, \ldots, b_n$ is $B = AA^T$ where A is the matrix $A_{ij} = \alpha_{ij}$. Consequently, if $b_1, \ldots, b_n$ is a basis of a lattice L, then* $\det L = |\det A|$.

Proof. The assertion follows from the equations $(b_i, b_j) = \sum_{k=1}^{n} \alpha_{ik} \alpha_{jk}$. ∎

1.4.2. Short lattice vectors

We will need a fundamental result in convex geometry. In order to prepare for this, we introduce some simple notation. Let $H \subseteq \mathbb{R}^n$. The set H is said to be ***centrally symmetric***, if $v \in H$ implies $-v \in H$. The set H is ***convex,*** if $u, v \in H$ implies $\lambda u + (1 - \lambda) v \in H$ for all $0 \le \lambda \le 1$.

Theorem 1.53 (Minkowski's Convex Body Theorem). *Let L be a lattice in $\mathbb{R}^n$ and let $K \subseteq \mathbb{R}^n$ be a centrally symmetric, bounded, closed, convex set. Suppose that the volume of K is at least $2^n \det L$. Then $K \cap L \neq \{0\}$.*

Proof. By the conditions, the volume of the set $(1/2)K := \{(1/2)v : \ v \in K\}$ is at least $\det L$. Let $b_1, \ldots, b_n$ be a basis of the lattice L and let $P = \{\sum_{i=1}^{n} \alpha_i b_i : \ 0 \le \alpha_1, \ldots, \alpha_n < 1\}$ be the corresponding half-open parallelepiped. Then each of the vectors in $\mathbb{R}^n$ can be written uniquely in the form $x + z$ where $x \in L$ and $z \in P$. For an arbitrary lattice vector $x \in L$, we let

$$K_x = (1/2)K \cap (x + P) = (1/2)K \cap \{x + z : \ z \in P\} \ .$$

As the sets $(1/2)K$ and P are bounded, so is the set

$$(1/2)K - P = \{u - v : \ u \in (1/2) \cdot K, \ v \in P\} .$$

As L is discrete, L only has finitely many points in $(1/2)K - P$; that is, $K_x = \emptyset$, except for finitely many $x \in L$. Hence $S = \{x \in L : \ K_x \neq \emptyset\}$ is a finite set, and, moreover, the set $(1/2)K$ is the disjoint union of the sets K_x $(x \in S)$. Therefore, the total volume of these sets

is at least $\det L$. For a given $x \in S$, we set $P_x = K_x - x = \{z \in P : \ x + z \in (1/2)K\}$. Consider the closure $\overline{P}$ and $\overline{P_x}$ of the sets P and P_x, respectively:

$$\overline{P} = \left\{ \sum_{i=1}^{n} \alpha_i b_i : \ 0 \le \alpha_1, \ldots, \alpha_n \le 1 \right\}$$

and $\overline{P_x} = \left\{ z \in \overline{P} : \ x + z \in (1/2)K \right\}$. The total volume of the closed sets $\overline{P_x} \subseteq \overline{P}$ is at least as large as the volume of the set $\overline{P}$, and so these sets cannot be disjoint: there are $x \neq y \in S$ and $z \in \overline{P}$ such that $z \in \overline{P_x} \cap \overline{P_y}$, that is, $x + z \in (1/2)K$ and $y + z \in (1/2)K$. As $(1/2) \cdot K$ is centrally symmetric, we find that $-y - z \in (1/2) \cdot K$. As $(1/2)K$ is convex, we also have $(x-y)/2 = ((x+z)+(-y-z))/2 \in (1/2)K$. Hence $x - y \in K$. On the other hand, the difference $x - y$ of two lattice points lies in $L \setminus \{0\}$. ■

Minkowski's theorem is sharp. For, let $\epsilon > 0$ be an arbitrary positive number, and let $L = \mathbf{Z}^n$ be the lattice of points with integer coordinates in $\mathbb{R}^n$. Let K be the set of vectors $(v_1, \ldots, v_n) \in \mathbb{R}^n$ for which $-1 + \epsilon \le v_i \le 1 - \epsilon$ $(i = 1, \ldots, n)$. Then K is bounded, closed, convex, centrally symmetric with respect to the origin, its volume is $(1 - \epsilon)^n 2^n \det L$, yet $L \cap K = \{0\}$.

Corollary 1.54 *Let L be a lattice in $\mathbb{R}^n$. Then L has a lattice vector $v \neq 0$ whose length is at most $\sqrt{n} \sqrt[n]{\det L}$.*

Proof. Let K be the following centrally symmetric cube with side length $s = 2\sqrt[n]{\det L}$:

$$K = \{(v_1, \ldots, v_n) \in \mathbb{R}^n : \ -s/2 \le v_i \le s/2, \ i = 1, \ldots, n\} .$$

The volume of the cube K is exactly $2^n \det L$, and so it contains a non-zero lattice vector. However, the vectors in K have length at most $\sqrt{n} \sqrt[n]{\det L}$. ■

We remark that, for $n > 1$, we can find an even shorter lattice vector, if we replace the cube in the proof of the previous assertion by a suitable ball.

1.4.3. Gauss' algorithm for two-dimensional lattices

Our goal is to design an algorithm that finds a non-zero short vector in a given lattice. In this section we consider this problem for two-dimensional lattices, which is the simplest non-trivial case. Then there is an elegant, instructive, and efficient algorithm that finds short lattice vectors. This algorithm also serves as a basis for the higher-dimensional cases. Let L be a lattice with basis b_1, b_2 in $\mathbb{R}^2$.

GAUSS(b_1, b_2)

```
1  (a, b) ← (b_1, b_2)
2  forever
3        do b ← the shortest lattice vector on the line b − λa
4        if |b| < |a|
5           then b ↔ a
6           else return (a, b)
```

In order to analyse the procedure, the following facts will be useful.

Lemma 1.55 *Suppose that a and b are two linearly independent vectors in the plane $\mathbb{R}^2$, and let L be the lattice generated by them. The vector b is a shortest non-zero vector of L on the line $b - \lambda a$ if and only if*

$$|(b,a)/(a,a)| \leq 1/2 \ . \tag{1.10}$$

Proof. We write b as the sum of a vector parallel to a and a vector orthogonal to a:

$$b = (b,a)/(a,a)a + b^* \ . \tag{1.11}$$

Then, as the vectors a and b^* are orthogonal,

$$|b - \lambda a|^2 = \left|\left(\frac{(b,a)}{(a,a)} - \lambda\right)a + b^*\right|^2 = \left(\frac{(b,a)}{(a,a)} - \lambda\right)^2 |a|^2 + |b^*|^2 \ .$$

This quantity takes its smallest value for the integer λ that is the closest to the number $(b,a)/(a,a)$. Hence $\lambda = 0$ gives the minimal value if and only if (1.10) holds. ∎

Lemma 1.56 *Suppose that the linearly independent vectors a and b form a basis for a lattice $L \subseteq \mathbb{R}^2$ and that inequality (1.10) holds. Assume, further, that*

$$|b|^2 \geq (3/4)|a|^2 \ . \tag{1.12}$$

Write b, as in (1.11), as the sum of the vector $((b,a)/(a,a))a$, which is parallel to a, and the vector $b^* = b - ((b,a)/(a,a))a$, which is orthogonal to a. Then

$$|b^*|^2 \geq (1/2)|a|^2 \ . \tag{1.13}$$

Further, either b or a is a shortest non-zero vector in L.

Proof. By the assumptions,

$$|a|^2 \leq \frac{4}{3}|b|^2 = \frac{4}{3}|b^*|^2 + \frac{4}{3}\left((b,a)/(a,a)\right)^2 |a|^2 \leq \frac{4}{3}|b^*|^2 + (1/3)|a|^2 \ .$$

Rearranging the last displayed line, we obtain $|b^*|^2 \geq (1/2)|a|^2$.

The length of a vector $0 \neq v = \alpha a + \beta b \in L$ can be computed as

$$|\alpha a + \beta b|^2 = |\beta b^*|^2 + (\alpha + \beta(b,a)/(a,a))^2 |a|^2 \geq \beta^2 |b^*|^2 \geq (1/2)\beta^2 |a|^2,$$

which implies $|v| > |a|$ whenever $|\beta| \geq 2$. If $\beta = 0$ and $\alpha \neq 0$, then $|v| = |\alpha| \cdot |a| \geq |a|$. Similarly, $\alpha = 0$ and $\beta \neq 0$ gives $|v| = |\beta| \cdot |b| \geq |b|$. It remains to consider the case when $\alpha \neq 0$ and $\beta = \pm 1$. As $|-v| = |v|$, we may assume that $\beta = 1$. In this case, however, v is of the form $v = b - \lambda a$ ($\lambda = -\alpha$), and, by Lemma 1.55, the vector b is a shortest lattice vector on this line. ∎

Theorem 1.57 *Let v be a shortest non-zero lattice vector in L. Then Gauss' algorithm terminates after $O(1 + \lg(|b_1|/|v|))$ iterations, and the resulting vector a is a shortest non-zero vector in L.*

Proof. First we verify that, during the course of the algorithm, the vectors a and b will always form a basis for the lattice L. If, in line 3, we replace b by a vector of the form $b' = b - \lambda a$, then, as $b = b' + \lambda a$, the pair $a,\ b'$ remains a basis of L. The swap in line 5 only concerns the order of the basis vectors. Thus a and b is always a basis of L, as we claimed.

By Lemma 1.55, inequality (1.10) holds after the first step (line 3) in the loop, and so we may apply Lemma 1.56 to the scenario before lines 4–5. This shows that if none of a and b is shortest, then $|b|^2 \leq (3/4)|a|^2$. Thus, except perhaps for the last execution of the loop, after each swap in line 5, the length of a is decreased by a factor of at least $\sqrt{3/4}$. Thus we obtain the bound for the number of executions of the loop. Lemma 1.56 implies also that the vector a at the end is a shortest non-zero vector in L. ■

Gauss' algorithm gives an efficient polynomial time method for computing a shortest vector in the lattice $L \subseteq \mathbb{R}^2$. The analysis of the algorithm gives the following interesting theoretical consequence.

Corollary 1.58 *Let L be a lattice in $\mathbb{R}^2$, and let a be a shortest non-zero lattice vector in L. Then $|a|^2 \leq (2/\sqrt{3}) \det L$.*

Proof. Let b be a vector in L such that b is linearly independent of a and (1.10) holds. Then

$$|a|^2 \leq |b|^2 = |b^*|^2 + \left(\frac{(b,a)}{(a,a)}\right)^2 |a|^2 \leq |b^*|^2 + \frac{1}{4}|a|^2 \ ,$$

which yields $(3/4)|a|^2 \leq |b^*|^2$. The area of the fundamental parallelogram can be computed using the well-known formula

$$\text{area} = \text{base} \cdot \text{height},$$

and so $\det L = |a||b^*|$. The number $|b^*|$ can now be bounded by the previous inequality. ■

1.4.4. A Gram-Schmidt orthogonalisation and weak reduction

Let $b_1, \ldots, b_n$ be a linearly independent collection of vectors in $\mathbb{R}^n$. For an index i with $i \in \{1, \ldots, n\}$, we let b_i^* denote the component of b_i that is orthogonal to the subspace spanned by $b_1, \ldots, b_{i-1}$. That is,

$$b_i = b_i^* + \sum_{j=1}^{i-1} \lambda_{ij} b_j \ ,$$

where

$$(b_i^*, b_j) = 0 \quad \text{for} \quad j = 1, \ldots, i-1 \ .$$

Clearly $b_1^* = b_1$. The vectors $b_1^*, \ldots, b_{i-1}^*$ span the same subspace as the vectors $b_1, \ldots, b_{i-1}$, and so, with suitable coefficients μ_{ij}, we may write

$$b_i = b_i^* + \sum_{j=1}^{i-1} \mu_{ij} b_j^* \ , \tag{1.14}$$

and

$$(b_i^*, b_j^*) = 0, \quad \text{if } j \neq i \ .$$

By the latter equations, the vectors $b_1^*, \ldots, b_{i-1}^*, b_i^*$ form an orthogonal system, and so

$$\mu_{ij} = \frac{(b_i, b_j^*)}{(b_j^*, b_j^*)} \quad (j = 1, \ldots, i-1) \; . \tag{1.15}$$

The set of the vectors $b_1^*, \ldots, b_n^*$ is said to be the ***Gram-Schmidt orthogonalisation*** of the vectors $b_1, \ldots, b_n$.

Lemma 1.59 *Let $L \subseteq \mathbb{R}^n$ be a lattice with basis $b_1, \ldots, b_n$. Then*

$$\det L = \prod_{i=1}^{n} |b_i^*| \; .$$

Proof. Set $\mu_{ii} = 1$ and $\mu_{ij} = 0$, if $j > i$. Then $b_i^* = \sum_{k=1}^{n} \mu_{ik} b_k$, and so

$$(b_i^*, b_j^*) = \sum_{k=1}^{n} \mu_{ik} \sum_{l=1}^{n} (b_k, b_l) \mu_{jl},$$

that is, $B^* = MBM^T$ where B and B^* are the Gram matrices of the collections $b_1, \ldots, b_n$ and $b_1^*, \ldots, b_n^*$, respectively, and M is the matrix with entries μ_{ij}. The matrix M is a lower triangular matrix with ones in the main diagonal, and so $\det M = \det M^T = 1$. As B^* is a diagonal matrix, we obtain $\prod_{i=1}^{n} |b_i^*|^2 = \det B^* = (\det M)(\det B)(\det M^T) = \det B$. ∎

Corollary 1.60 (Hadamard inequality). $\prod_{i=1}^{n} |b_i| \geq \det L$.

Proof. The vector b_i can be written as the sum of the vector b_i^* and a vector orthogonal to b_i^*, and hence $|b_i^*| \leq |b_i|$. ∎

The vector b_i^* is the component of b_i orthogonal to the subspace spanned by the vectors $b_1, \ldots, b_{i-1}$. Thus b_i^* does not change if we subtract a linear combination of the vectors $b_1, \ldots, b_{i-1}$ from b_i. If, in this linear combination, the coefficients are integers, then the new sequence $b_1, \ldots, b_n$ will be a basis of the same lattice as the original. Similarly to the first step of the loop in Gauss' algorithm, we can make the numbers μ_{ij} in (1.15) small. The input of the following procedure is a basis $b_1, \ldots, b_n$ of a lattice L.

WEAK-REDUCTION$(b_1, \ldots, b_n)$

1 **for** $j \leftarrow n-1$ **downto** 1
2 **do for** $i \leftarrow j+1$ **to** n
3 $b_i \leftarrow b_i - \lambda b_j$, where λ is the integer nearest the number $(b_i, b_j^*)/(b_j^*, b_j^*)$
4 **return** $(b_1, \ldots, b_n)$

Definition 1.61 (Weakly reduced basis). *A basis $b_1, \ldots, b_n$ of a lattice is said to be* ***weakly reduced*** *if the coefficients μ_{ij} in (1.15) satisfy*

$$|\mu_{ij}| \leq \frac{1}{2} \quad \textit{for} \quad 1 \leq j < i \leq n \; .$$

Lemma 1.62 *The basis given by the procedure* WEAK-REDUCTION *is weakly reduced.*

Proof. By the remark preceding the algorithm, we obtain that the vectors $b_1^*, \ldots, b_n^*$ never change. Indeed, we only subtract linear combinations of vectors with index less than i from b_i. Hence the inner instruction does not change the value of (b_k, b_l^*) with $k \neq i$. The values of the (b_i, b_l^*) do not change for $l > j$ either. On the other hand, the instruction achieves, with the new b_i, that the inequality $|\mu_{ij}| \leq 1/2$ holds:

$$|(b_i - \lambda b_j^*, b_j^*)| = |(b_i, b_j^*) - \lambda(b_j^*, b_j^*)| = |(b_i, b_j^*) - \lambda(b_j^*, b_j^*)| \leq \frac{1}{2}(b_j^*, b_j^*) \ .$$

By the observations above, this inequality remains valid during the execution of the procedure. ■

1.4.5. Lovász-reduction

First we define, in an arbitrary dimension, a property of the bases that usually turns out to be useful. The definition will be of a technical nature. Later we will see that these bases are interesting, in the sense that they consist of short vectors. This property will make them widely applicable.

Definition 1.63 *A basis $b_1, \ldots, b_n$ of a lattice L is said to be **(Lovász-)reduced** if*

- *it is weakly reduced,*

and, using the notation introduced for the Gram-Schmidt orthogonalisation,

- $|b_i^*|^2 \leq (/3)|b_{i+1}^* + \mu_{i+1,i} b_i^*|^2$ *for all* $1 \leq i < n$.

Let us observe the analogy of the conditions above to the inequalities that we have seen when investigating Gauss' algorithm. For $i = 1$, $a = b_1$ and $b = b_2$, being weakly reduced ensures that b is a shortest vector on the line $b - \lambda a$. The second condition is equivalent to the inequality $|b|^2 \geq (3/4)|a|^2$, but here it is expressed in terms of the Gram-Schmidt basis. For a general index i, the same is true, if a plays the rôle of the vector b_i, and b plays the rôle of the component of the vector b_{i+1} that is orthogonal to the subspace spanned by $b_1, \ldots, b_{i-1}$.

Lovász-Reduction$(b_1, \ldots, b_n)$

```
1  forever
2        do (b_1, ..., b_n) ←Weak-Reduction(b_1, ..., b_n)
3        find an index i for which the second condition of being reduced is violated
4        if there is such an i
5           then b_i ↔ b_{i+1}
6           else return (b_1, ..., b_n)
```

Theorem 1.64 *Suppose that in the lattice $L \subseteq \mathbb{R}^n$ each of the pairs of the lattice vectors has an integer scalar product. Then the swap in the 5th line of the* Lovász-Reduction *occurs at most $\lg_{4/3}(B_1 \cdots B_{n-1})$ times where B_i is the upper left $(i \times i)$-subdeterminant of the Gram matrix of the initial basis $b_1, \ldots, b_n$.*

Proof. The determinant B_i is the determinant of the Gram matrix of $b_1, \ldots, b_i$, and, by the observations we made at the discussion of the Gram-Schmidt orthogonalisation, $B_i =$

$\prod_{j=1}^{i} |b_j^*|^2$. This, of course, implies that $B_i = B_{i-1}|b_i^*|^2$ for $i > 1$. By the above, the procedure Weak-Reduction cannot change the vectors b_i^*, and so it does not change the product $\prod_{j=1}^{n-1} B_j$ either. Assume, in line 5 of the procedure, that a swap $b_i \leftrightarrow b_{i+1}$ takes place. Observe that, unless $j = i$, the sets $\{b_1, \ldots, b_j\}$ do not change, and neither do the determinants B_j. The rôle of the vector b_i^* is taken over by the vector $b_{i+1}^* + \mu_{i,i+1} b_i$, whose length, because of the conditions of the swap, is at most $\sqrt{3/4}$ times the length of b_i^*. That is, the new B_i is at most 3/4 times the old. By the observation above, the new value of $B = \prod_{j=1}^{n-1} B_j$ will also be at most 3/4 times the old one. Then the assertion follows from the fact that the quantity B remains a positive integer. ∎

Corollary 1.65 *Under the conditions of the previous theorem, the cost of the procedure* Lovász-Reduction *is at most* $O(n^5 \lg nC)$ *arithmetic operations with rational numbers where* C *is the maximum of* 2 *and the quantities* $|(b_i, b_j)|$ *with* $i, j = 1, \ldots, n$.

Proof. It follows from the Hadamard inequality that

$$B_i \le \prod_{j=1}^{i} \sqrt{(b_1, b_j)^2 + \ldots + (b_i, b_j)^2} \le (\sqrt{i}C)^i \le (\sqrt{n}C)^n \ .$$

Hence $B_1 \cdots B_{n-1} \le (\sqrt{n}C)^{n(n-1)}$ and $\lg_{4/3}(B_1 \ldots B_{n-1}) = O(n^2 \lg nC)$. By the previous theorem, this is the number of iterations in the algorithm. The cost of the Gram–Schmidt orthogonalisation is $O(n^3)$ operations, and the cost of weak reduction is $O(n^2)$ scalar product computations, each of which can be performed using $O(n)$ operations (provided the vectors are represented by their coordinates in an orthogonal basis). ∎

One can show that the length of the integers that occur during the run of the algorithm (including the numerators and the denominators of the fractions in the Gram–Schmidt orthogonalisation) will be below a polynomial bound.

1.4.6. Properties of reduced bases

Theorem 1.67 of this section gives a summary of the properties of reduced bases that turn out to be useful in their applications. We will find that a reduced basis consists of relatively short vectors. More precisely, $|b_1|$ will approximate, within a constant factor depending only on the dimension, the length of a shortest non-zero lattice vector.

Lemma 1.66 *Let us assume that the vectors* $b_1, \ldots, b_n$ *form a reduced basis of a lattice* L. *Then, for* $1 \le j \le i \le n$,

$$(b_i^*, b_i^*) \ge 2^{j-i} (b_j^*, b_j^*) \ . \tag{1.16}$$

In particular,

$$(b_i^*, b_i^*) \ge 2^{1-i} (b_1^*, b_1^*) \ . \tag{1.17}$$

Proof. Substituting $a = b_i^*$, $b = b_{i+1}^* + ((b_{i+1}, b_i^*))/((b_i^*, b_i^*)b_i^*)$, Lemma 1.56 gives, for all $1 \le i < n$, that

$$(b_{i+1}^*, b_{i+1}^*) \ge (1/2)(b_i^*, b_i^*) \ .$$

Thus, inequality (1.16) follows by induction. ∎

Now we can formulate the fundamental theorem of reduced bases.

Theorem 1.67 *Assume that the vectors $b_1, \ldots, b_n$ form a reduced basis of a lattice L. Then*

(i) $|b_1| \le 2^{(n-1)/4} (\det L)^{(1/n)}$.

(ii) $|b_1| \le 2^{(n-1)/2} |b|$ *for all lattice vectors* $0 \ne b \in L$. *In particular, the length of* b_1 *is not greater than* $2^{(n-1)/2}$ *times the length of a shortest non-zero lattice vector.*

(iii) $|b_1| \cdots |b_n| \le 2^{(n(n-1))/4} \det L$.

Proof. (i) Using inequality (1.17),

$$(\det L)^2 = \prod_{i=1}^{n} (b_i^*, b_i^*) \ge \prod_{i=1}^{n} (2^{1-i}(b_1, b_1)) = 2^{\frac{-n(n-1)}{2}} (b_1, b_1)^n \ ,$$

and so assertion (i) holds.

(ii) Let $b = \sum_{i=1}^{n} z_i b_i \in L$ with $z_i \in \mathbf{Z}$ be a lattice vector. Assume that z_j is the last non-zero coefficient and write $b_j = b_j^* + v$ where v is a linear combination of the vectors $b_1, \ldots, b_{j-1}$. Hence $b = z_j b_j^* + w$ where w lies in the subspace spanned by $b_1, \ldots, b_{j-1}$. As b_j^* is orthogonal to this subspace,

$$(b, b) = z_j^2 (b_j^*, b_j^*) + (w, w) \ge (b_j^*, b_j^*) \ge 2^{1-j} (b_1, b_1) \ge 2^{1-n} (b_1, b_1) \ ,$$

and so assertion (ii) is valid.

(iii) First we show that $(b_i, b_i) \le 2^{i-1} (b_i^*, b_i^*)$. This inequality is obvious if $i = 1$, and so we assume that $i > 1$. Using the decomposition (1.14) of the vector b_i and the fact that the basis is weakly reduced, we obtain that

$$\begin{aligned} (b_i, b_i) &= \sum_{j=1}^{i} \left(\frac{(b_i, b_j^*)}{(b_j^*, b_j^*)} \right)^2 (b_j^*, b_j^*) \le (b_i^*, b_i^*) + \frac{1}{4} \sum_{j=1}^{i-1} (b_j^*, b_j^*) \le (b_i^*, b_i^*) + \frac{1}{4} \sum_{j=1}^{i-1} 2^{i-j} (b_i^*, b_i^*) \\ &\le (2^{i-2} + 1)(b_i^*, b_i^*) \le 2^{i-1} (b_i^*, b_i^*) \ . \end{aligned}$$

Multiplying these inequalities for $i = 1, \ldots, n$,

$$\prod_{i=1}^{n} (b_i, b_i) \le \prod_{i=1}^{n} 2^{i-1} (b_i^*, b_i^*) = 2^{\frac{n(n-1)}{2}} \prod_{i=1}^{n} (b_i^*, b_i^*) = 2^{\frac{n(n-1)}{2}} (\det L)^2 \ ,$$

which is precisely the inequality in (iii). ∎

It is interesting to compare assertion (i) in the previous theorem and Corollary 1.54 after Minkowski's theorem. Here we obtain a weaker bound for the length of b_1, but this vector can be obtained by an efficient algorithm. Essentially, the existence of the basis that satisfies assertion (iii) was first shown by Hermite using the tools in the proofs of Theorems 1.48 and 1.67. Using a Lovász-reduced basis, the cost of finding a shortest vector in a lattice with dimension n is at most polynomial in the input size and in 3^{n^2}; see Exercise 1.4-4..

Exercises

1.4-1 *The triangular lattice is optimal.* Show that the bound in Corollary 1.58 is sharp. More precisely, let $L \subseteq \mathbb{R}^2$ be a full lattice and let $0 \ne a \in L$ be a shortest vector in L. Verify

that the inequality $|a|^2 = (2/\sqrt{3})\det L$ holds if and only if L is similar to the triangular lattice.

1.4-2 *The denominators of the Gram-Schmidt numbers.* Let us assume that the Gram matrix of a basis $b_1, \ldots, b_n$ has only integer entries. Show that the numbers μ_{ij} in (1.15) can be written in the form $\mu_{ij} = \zeta_{ij}/\prod_{k=1}^{j-1} B_k$ where the ζ_{ij} are integers and B_k is the determinant of the Gram matrix of the vectors $b_1, \ldots, b_k$.

1.4-3 *The length of the vectors in a reduced basis.* Let $b_1, \ldots, b_n$ be a reduced basis of a lattice L and let us assume that the numbers (b_i, b_i) are integers. Give an upper bound depending only on n and $\det L$ for the length of the vectors b_i. More precisely, prove that

$$|b_i| \le 2^{\frac{n(n-1)}{4}} \det L \,.$$

1.4-4 *The coordinates of a shortest lattice vector.* Let $b_1, \ldots, b_n$ be a reduced basis of a lattice L. Show that each of the shortest vectors in L is of the form $\sum z_i b_i$ where $z_i \in \mathbf{Z}$ and $|z_i| \le 3^n$. Consequently, for a bounded n, one can find a shortest non-zero lattice vector in polynomial time.

Hint: Assume, for some lattice vector $v = \sum z_i b_i$, that $|v| \le |b_1|$. Let us write v in the basis $b_1^*, \ldots, b_n^*$:

$$v = \sum_{j=1}^{n} (z_j + \sum_{i=j+1}^{n} \mu_{ij} z_i) b_j^* \,.$$

It follows from the assumption that each of the components of v (in the orthogonal basis) is at most as long as $b_1 = b_1^*$:

$$\left| z_j + \sum_{i=j+1}^{n} \mu_{ij} z_i \right| \le \frac{|b_1^*|}{|b_j^*|} \,.$$

Use then the inequalities $|\mu_{ij}| \le 1/2$ and (1.17).

1.5. Factoring polynomials in $\mathbb{Q}[x]$

In this section we study the problem of factoring polynomials with rational coefficients. The input of the ***factorisation problem*** is a polynomial $f(x) \in \mathbb{Q}[x]$. Our goal is to compute a factorisation

$$f = f_1^{e_1} f_2^{e_2} \cdots f_s^{e_s}, \tag{1.18}$$

where the polynomials $f_1, \ldots, f_s$ are pairwise relatively prime, and irreducible over $\mathbb{Q}$, and the numbers e_i are positive integers. By Theorem 1.4, f determines, essentially uniquely, the polynomials f_i and the exponents e_i.

1.5.1. Preparations

First we reduce the problem (1.18) to another problem that can be handled more easily.

Lemma 1.68 *We may assume that the polynomial $f(x)$ has integer coefficients and it has leading coefficient* 1.

Proof. Multiplying by the common denominator of the coefficients, we may assume that $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbf{Z}[x]$. Performing the substitution $y = a_nx$, we obtain the polynomial

$$g(y) = {a_n}^{n-1} f\left(\frac{y}{a_n}\right) = y^n + \sum_{i=0}^{n-1} a_n^{n-i-1} a_i y^i,$$

which has integer coefficients and its leading coefficient is 1. Using a factorisation of $g(y)$, a factorisation of $f(x)$ can be obtained efficiently. ∎

Primitive polynomials, Gauss' lemma

Definition 1.69 *A polynomial $f(x) \in \mathbf{Z}[x]$ is said to be **primitive**, if the greatest common divisor of its coefficients is* 1.

A polynomial $f(x) \in \mathbf{Z}[x] \setminus \{0\}$ can be written in a unique way as the product of an integer and a primitive polynomial in $\mathbf{Z}[x]$. Indeed, if a is the greatest common divisor of the coefficients, then $f(x) = a(1/a)f(x)$. Clearly, $(1/a)f(x)$ is a primitive polynomial with integer coefficients.

Lemma 1.70 (Gauss' Lemma). *If $u(x)$, $v(x) \in \mathbf{Z}[x]$ are primitive polynomials, then so is the product $u(x)v(x)$.*

Proof. We argue by contradiction and assume that p is a prime number that divides all the coefficients of uv. Set $u(x) = \sum_{i=0}^{n} u_i x^i$, $v(x) = \sum_{j=0}^{m} v_j x^j$ and let i_0 and j_0 be the smallest indices such that $p \nmid u_{i_0}$ and $p \nmid v_{j_0}$. Let $k_0 = i_0 + j_0$ and consider the coefficient of x^{k_0} in the product $u(x)v(x)$. This coefficient is

$$\sum_{i+j=k_0} u_i v_j = u_{i_0} v_{j_0} + \sum_{i=0}^{i_0-1} u_i v_{k_0-i} + \sum_{j=0}^{j_0-1} u_{k_0-j} v_j \ .$$

Both of the sums on the right-hand side of this equation are divisible by p, while $u_{i_0}v_{j_0}$ is not, and hence the coefficient of x^{k_0} in $u(x)v(x)$ cannot be divisible by p after all. This, however, is a contradiction. ∎

Proposition 1.71 *Let us assume that $g(x)$, $h(x) \in \mathbb{Q}[x]$ are polynomials with rational coefficients and leading coefficient* 1 *such that the product $g(x)h(x)$ has integer coefficients. Then the polynomials $g(x)$ and $h(x)$ have integer coefficients.*

Proof. Let us multiply $g(x)$ and $h(x)$ by the least common multiple c_g and c_h, respectively, of the denominators of their coefficients. Then the polynomials $c_g g(x)$ and $c_h h(x)$ are primitive polynomials with integer coefficients. Hence, by Gauss' Lemma, so is the product $c_g c_h g(x)h(x) = (c_g g(x))(c_h h(x))$. As the coefficients of $g(x)h(x)$ are integers, each of its coefficients is divisible by the integer $c_g c_h$. Hence $c_g c_h = 1$, and so $c_g = c_h = 1$. Therefore $g(x)$ and $h(x)$ are indeed polynomials with integer coefficients. ∎

One can show similarly, for a polynomial $f(x) \in \mathbf{Z}[x]$, that factoring $f(x)$ in $\mathbf{Z}[x]$ is equivalent to factoring the primitive part of $f(x)$ in $\mathbb{Q}[x]$ and factoring an integer, namely the greatest common divisor of the coefficients

Mignotte's bound

As we work over an infinite field, we have to pay attention to the size of the results in our computations.

Definition 1.72 *The **norm** of a polynomial $f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{C}[x]$ with complex coefficients is the real number $\|f(x)\| = \sqrt{\sum_{i=0}^{n} |a_i|^2}$.*

The inequality $\max_{i=0}^{n} |a_i| \leq \|f(x)\|$ implies that a polynomial $f(x)$ with integer coefficients can be described using $O(n \lg \|f(x)\|)$ bits.

Lemma 1.73 *Let $f(x) \in \mathbb{C}[x]$ be a polynomial with complex coefficients. Then, for all $c \in \mathbb{C}$, we have*

$$\|(x - c)f(x)\| = \|(\overline{c}x - 1)f(x)\| ,$$

where $\overline{c}$ is the usual conjugate of the complex number c.

Proof. Let us assume that $f(x) = \sum_{i=0}^{n} a_i x^i$ and set $a_{n+1} = a_{-1} = 0$. Then

$$(x - c)f(x) = \sum_{i=0}^{n+1} (a_{i-1} - ca_i)x^i ,$$

and hence

$$\begin{aligned} \|(x - c)f(x)\|^2 &= \sum_{i=0}^{n+1} |a_{i-1} - ca_i|^2 = \sum_{i=0}^{n+1} (|a_{i-1}|^2 + |ca_i|^2 - a_{i-1}\overline{ca_i} - \overline{a_{i-1}}ca_i) \\ &= \|f(x)\|^2 + |c|^2\|f(x)\|^2 - \sum_{i=0}^{n+1} (a_{i-1}\overline{ca_i} + \overline{a_{i-1}}ca_i) . \end{aligned}$$

Performing similar computations with the right-hand side of the equation in the lemma, we obtain that

$$(\overline{c}x - 1)f(x) = \sum_{i=0}^{n+1} (\overline{c}a_{i-1} - a_i)x^i ,$$

and so

$$\begin{aligned} \|(\overline{c}x - 1)f(x)\|^2 &= \sum_{i=0}^{n+1} |\overline{c}a_{i-1} - a_i|^2 = \sum_{i=0}^{n+1} (|\overline{c}a_{i-1}|^2 + |a_i|^2 - \overline{c}a_{i-1}\overline{a_i} - c\overline{a_{i-1}}a_i) \\ &= \|f(x)\|^2 + |c|^2\|f(x)\|^2 - \sum_{i=0}^{n+1} (a_{i-1}\overline{ca_i} + \overline{a_{i-1}}ca_i) . \end{aligned}$$

The proof of the lemma is now complete. ∎

Theorem 1.74 (Mignotte). *Let us assume that the polynomials $f(x), g(x) \in \mathbb{C}[x]$ have complex coefficients and leading coefficient 1 and that $g(x)|f(x)$. If $\deg(g(x)) = m$, then $\|g(x)\| \leq 2^m \|f(x)\|$.*

Proof. By the fundamental theorem of algebra, $f(x) = \prod_{i=1}^{n}(x-\alpha_i)$ where $\alpha_1, \ldots, \alpha_n$ are the complex roots of the polynomial $f(x)$ (with multiplicity). Then there is a subset $I \subseteq \{1, \ldots, n\}$ such that $g(x) = \prod_{i\in I}(x-\alpha_i)$. First we claim, for an arbitrary set $J \subseteq \{1, \ldots, n\}$, that

$$\prod_{i\in J} |\alpha_i| \leq \|f(x)\| \,. \tag{1.19}$$

If J contains an integer i with $\alpha_i = 0$, then this inequality will trivially hold. Let us hence assume that $\alpha_i \neq 0$ for every $i \in J$. Set $\overline{J} = \{1, \ldots, n\} \setminus J$ and $h(x) = \prod_{i\in\overline{J}}(x-\alpha_i)$. Applying Lemma 1.73 several times, we obtain that

$$\|f(x)\| = \|\prod_{i\in J}(x-\alpha_i)h(x)\| = \|\prod_{i\in J}(\overline{\alpha_i}x-1)h(x)\| = |\prod_{i\in J}\overline{\alpha_i}| \cdot \|u(x)\| \,,$$

where $u(x) = \prod_{i\in J}(x - 1/\overline{\alpha_i})h(x)$. As the leading coefficient of $u(x)$ is 1, $\|u(x)\| \geq 1$, and so

$$|\prod_{i\in J}\alpha_i| = |\prod_{i\in J}\overline{\alpha_i}| = \|f(x)\|/\|u(x)\| \leq \|f(x)\| \,.$$

Let us express the coefficients of $g(x)$ using its roots:

$$\begin{aligned} g(x) &= \prod_{i\in I}(x-\alpha_i) = \sum_{J\subseteq I}\left((-1)^{|J|}\prod_{j\in J}\alpha_j x^{m-|J|}\right) \\ &= \sum_{i=0}^{m}(-1)^{m-i}\left(\sum_{J\subseteq I,|J|=m-i}\prod_{j\in J}\alpha_j\right)x^i \,. \end{aligned}$$

For an arbitrary polynomial $t(x) = t_0 + \cdots + t_k x^k$, the inequality $\|t(x)\| \leq |t_0| + \cdots + |t_k|$ is valid. Therefore, using inequality (1.19), we find that

$$\begin{aligned} \|g(x)\| &\leq \sum_{i=0}^{m}\left|\sum_{J\subseteq I,|J|=m-i}\prod_{j\in J}\alpha_j\right| \\ &\leq \sum_{J\subseteq I}\left|\prod_{j\in J}\alpha_j\right| \leq 2^m\|f(x)\| \,. \end{aligned}$$

The proof is now complete. ∎

Corollary 1.75 *The bit size of the irreducible factors in* $\mathbb{Q}[x]$ *of an* $f(x) \in \mathbf{Z}[x]$ *with leading coefficient* 1 *is polynomial in the bit size of* $f(x)$.

Resultant and good reduction

Let $\mathbb{F}$ be an arbitrary field, and let $f(x),\ g(x) \in \mathbb{F}[x]$ be polynomials with degree n and m, respectively: $f = a_0 + a_1 x + \ldots + a_n x^n$, $g = b_0 + b_1 x + \ldots + b_m x^m$ where $a_n \neq 0 \neq b_m$. We recall the concept of the ***resultant*** from Chapter **??**. The resultant of f and g is the determinant of

the $((m+n)\times(m+n))$-matrix

$$M = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots & a_n & & & \\ & a_0 & a_1 & a_2 & \cdots & a_{n-1} & a_n & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} & a_n \\ b_0 & b_1 & \cdots & b_{m-1} & b_m & & & & \\ & b_0 & b_1 & \cdots & b_{m-1} & b_m & & & \\ & & b_0 & b_1 & \cdots & b_{m-1} & b_m & & \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & & b_0 & b_1 & \cdots & b_{m-1} & b_m. \end{pmatrix}. \tag{1.20}$$

The matrix above is usually referred to as the Sylvester matrix. The blank spaces in the Sylvester matrix represent zero entries.

The resultant provides information about the common factors of f and g. One can use it to express, particularly elegantly, the fact that two polynomials are relatively prime:

$$\gcd(f(x), g(x)) = 1 \Leftrightarrow Res(f,g) \neq 0\ . \tag{1.21}$$

Corollary 1.76 *Let $f(x) = a_0 + a_1x + \cdots + a_nx^n \in \mathbf{Z}[x]$ be a square-free (in $\mathbb{Q}[x]$), non-constant polynomial. Then $Res(f(x), f'(x))$ is an integer. Further, assume that p is a prime not dividing na_n. Then the polynomial $f(x) \pmod p$ is square-free in $\mathbb{F}_p[x]$ if and only if p does not divide $Res(f(x), f'(x))$.*

Proof. The entries of the Sylvester matrix corresponding to $f(x)$ and $f'(x)$ are integers, and so is its determinant. The polynomial f has no multiple roots over $\mathbb{Q}$, and so, by Exercise 1.5-1., $\gcd(f(x), f'(x)) = 1$, which gives, using (1.21), that $Res(f(x), f'(x)) \neq 0$. Let $F(x)$ denote the polynomial f reduced modulo p. Then it follows from our assumptions that $Res(F(x), F'(x))$ is precisely the residue of $Res(f(x), f'(x))$ modulo p. By Exercise 1.5-1., the polynomial $F(x)$ is square-free precisely when $\gcd(F(x), F'(x)) = 1$, which is equivalent to $Res(F(x), F'(x)) \neq 0$. This amounts to saying that p does not divide the integer $Res(f(x), f'(x))$. ■

Corollary 1.77 *If $f(x) \in \mathbf{Z}[x]$ is a square-free polynomial with degree n, then there is a prime $p = O((n \lg n + 2n \lg \|f\|)^2)$ (that is, the absolute value of p is polynomial in the bit size of f) such that the polynomial $f(x) \pmod p$ is square-free in $\mathbb{F}_p[x]$.*

Proof. By the Prime Number Theorem (Theorem 33.37), for large enough K, the product of the primes in the interval $[1, K]$ is at least $2^{(0.9K/\ln K)}$. Reference to NA!

Set $K = ((n+1)\lg n + 2n \lg \|f\|)^2$. If K is large enough, then

$$p_1 \cdots p_l \geq 2^{(0.9K/\ln K)} > 2^{\sqrt{K}} \geq n^{n+1}\|f\|^{2n} \geq n^{n+1}\|f\|^{2n-1}|a_n| \tag{1.22}$$

where $p_1, \ldots, p_l$ are primes not larger than K, and a_n is the leading coefficient of f.

Let us suppose, for the primes $p_1, \ldots, p_l$, that $f(x) \pmod{p_i}$ is not square-free in $\mathbb{F}_{p_i}[x]$. Then the product $p_1 \cdots p_l$ divides $Res(f(x), f'(x)) \cdot na_n$, and so

$$p_1 \cdots p_l \leq |Res(f, f')| \cdot |na_n| \leq \|f\|^{n-1} \cdot \|f'\|^n \cdot |na_n| \leq n^{n+1}\|f\|^{2n-1}|a_n|\ .$$

(In the last two inequalities, we used the Hadamard inequality, and the fact that $\|f'(x)\| \leq n\|f(x)\|$.) This contradicts to inequality (1.22), which must be valid because of the choice of K. ■

We note that using the Prime Number Theorem more carefully, one can obtain a stronger bound for p.

Hensel lifting

We present a general procedure that can be used to obtain, given a factorisation modulo a prime p, a factorisation modulo p^N of a polynomial with integer coefficients.

Theorem 1.78 (Hensel's lemma). *Suppose that $f(x)$, $g(x)$, $h(x) \in \mathbf{Z}[x]$ are polynomials with leading coefficient 1 such that $f(x) \equiv g(x)h(x) \pmod{p}$, and, in addition, $g(x) \pmod{p}$ and $h(x) \pmod{p}$ are relatively prime in $\mathbb{F}_p[x]$. Then, for an arbitrary positive integer t, there are polynomials $g_t(x)$, $h_t(x) \in \mathbf{Z}[x]$ such that*

- *both of the leading coefficients of $g_t(x)$ and $h_t(x)$ are equal to 1,*
- *$g_t(x) \equiv g(x) \pmod{p}$ and $h_t(x) \equiv h(x) \pmod{p}$,*
- *$f(x) \equiv g_t(x)h_t(x) \pmod{p^t}$.*

Moreover, the polynomials $g_t(x)$ and $h_t(x)$ satisfying the conditions above are unique modulo p^t.

Proof. From the conditions concerning the leading coefficients, we obtain that $\deg f(x) = \deg g(x) + \deg h(x)$, and, further, that $\deg g_t(x) = \deg g(x)$ and $\deg h_t(x) = \deg h(x)$, provided the suitable polynomials $g_t(x)$ and $h_t(x)$ indeed exist. The existence is proved by induction on t. In the initial step, $t = 1$ and the choice $g_1(x) = g(x)$ and $h_1(x) = h(x)$ is as required.

The induction step $t \to t+1$: let us assume that there exist polynomials $g_t(x)$ and $h_t(x)$ that are well-defined modulo p^t and satisfy the conditions. If the polynomials $g_{t+1}(x)$ and $h_{t+1}(x)$ exist, then they must satisfy the conditions imposed on $g_t(x)$ and $h_t(x)$. As $g_t(x)$ and $h_t(x)$ are unique modulo p^t, we may write $g_{t+1}(x) = g_t(x) + p^t\delta_g(x)$ and $h_{t+1}(x) = h_t(x) + p^t\delta_h(x)$ where $\delta_g(x)$ and $\delta_h(x)$ are polynomials with integer coefficients. The condition concerning the leading coefficients guarantees that $\deg \delta_g(x) < \deg g(x)$ and that $\deg \delta_h(x) < \deg h(x)$.

By the induction hypothesis, $f(x) = g_t(x)h_t(x) + p^t\lambda(x)$ where $\lambda(x) \in \mathbf{Z}[x]$. The observations about the degrees of the polynomials $g_t(x)$ and $h_t(x)$ imply that the degree of $\lambda(x)$ is smaller than $\deg f(x)$. Now we may compute that

$$\begin{aligned} g_{t+1}(x)h_{t+1}(x) - f(x) &= g_t(x)h_t(x) - f(x) + p^t h_t(x)\delta_g(x) + p^t g_t(x)\delta_h(x) + p^{2t}\delta_g(x)\delta_h(x) \\ &\equiv -p^t\lambda(x) + p^t h_t(x)\delta_g(x) + p^t g_t(x)\delta_h(x) \pmod{p^{2t}} . \end{aligned}$$

As $2t > t+1$, the congruence above holds modulo p^{t+1}. Thus $g_{t+1}(x)$ and $h_{t+1}(x)$ satisfy the conditions if and only if

$$p^t h_t(x)\delta_g(x) + p^t g_t(x)\delta_h(x) \equiv p^t\lambda(x) \pmod{p^{t+1}} .$$

This, however, amounts to saying, after cancelling p^t from both sides, that

$$h_t(x)\delta_g(x) + g_t(x)\delta_h(x) \equiv \lambda(x) \pmod{p}.$$

Using the congruences $g_t(x) \equiv g(x) \pmod{p}$ and $h_t(x) \equiv h(x) \pmod{p}$ we obtain that this is equivalent to the congruence

$$h(x)\delta_g(x) + g(x)\delta_h(x) \equiv \lambda(x) \pmod{p}. \tag{1.23}$$

Considering the inequalities $\deg \delta_g(x) < \deg g_t(x)$ and $\deg \delta_h(x) < \deg h_t(x)$ and the fact that in $\mathbb{F}_p[x]$ the polynomials $g(x) \pmod{p}$ and $h(x) \pmod{p}$ are relatively prime, we find that equation (1.23) can be solved uniquely in $\mathbb{F}_p[x]$. For, if $u(x)$ and $v(x)$ form a solution to $u(x)g(x) + v(x)h(x) \equiv 1 \pmod{p}$, then, by Theorem 1.12, the polynomials

$$\delta_g(x) \quad = \quad v(x)\lambda(x) \pmod{g(x)},$$

and

$$\delta_h(x) \quad = \quad u(x)\lambda(x) \pmod{h(x)}$$

form a solution of (1.23). The uniqueness of the solution follows from the bounds on the degrees, and from the fact that $g(x) \pmod{p}$ and $h(x) \pmod{p}$ relatively prime. The details of this are left to the reader. ■

Corollary 1.79 *Assume that p, and the polynomials* $f(x)$, $g(x)$, $h(x) \in \mathbf{Z}[x]$ *satisfy the conditions of Hensel's lemma. Set* $\deg f = n$ *and let N be a positive integer. Then the polynomials* $g_N(x)$ *and* $h_N(x)$ *can be obtained using* $O(Nn^2)$ *arithmetic operations modulo* p^N.

Proof. The proof of Theorem 1.78 suggests the following algorithm.

HENSEL-LIFTING (f, g, h, p, N)

```
1  (u(x), v(x)) ← is a solution, in F_p[x], of u(x)g(x) + v(x)h(x) ≡ 1 (mod p)
2  (G(x), H(x)) ← (g(x), h(x))
3  for t ← 1 to N − 1
4      do λ(x) ← (f(x) − G(x) · H(x))/p^t
5         δ_g(x) ← v(x)λ(x) reduced modulo g(x) (in F_p[x])
6         δ_h(x) ← u(x)λ(x) reduced modulo h(x) (in F_p[x])
7         (G(x), H(x)) ← (G(x) + p^t δ_g(x), H(x) + p^t δ_h(x)) (in (Z/(p^{t+1}))[x])
8  return (G(x), H(x))
```

The polynomials u and v can be obtained using $O(n^2)$ operations in $\mathbb{F}_p$ (see Theorem 1.12 and the remark following it). An iteration $t \to t + 1$ consists of a constant number of operations with polynomials, and the cost of one run of the main loop is $O(n^2)$ operations (modulo p and p^{t+1}). The total cost of reaching $t = N$ is $O(Nn^2)$ operations. ■

1.5.2. The Berlekamp-Zassenhaus algorithm

The factorisation problem (1.18) was efficiently reduced to the case in which the polynomial f has integer coefficients and leading coefficient 1. We may also assume that $f(x)$ has no multiple factors in $\mathbb{Q}[x]$. Indeed, in our case $f'(x) \neq 0$, and so the possible multiple factors of f can be separated using the idea that we already used over finite fields as follows. By

Lemma 1.13, the polynomial $g(x) = f(x)/(f(x), f'(x))$ is already square-free, and, using Lemma 1.14, it suffices to find its factors with multiplicity one. From Proposition 1.71, we can see that $g(x)$ has integer coefficients and leading coefficient 1. Computing the greatest common divisor and dividing polynomials can be performed efficiently, and so the reduction can be carried out in polynomial time. (In the computation of the greatest common divisor, the intermediate expression swell can be avoided using the techniques presented in Chapter **??**.)

In the sequel we assume that the polynomial

$$f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in \mathbf{Z}[x]$$

we want to factor is square-free, its coefficients are integers, and its leading coefficient is 1.

The fundamental idea of the Berlekamp-Zassenhaus algorithm is that we compute the irreducible factors of $f(x)$ modulo p^N where p is a suitably chosen prime and N is large enough. If, for instance, $p^N > 2 \cdot 2^{n-1}\|f\|$, and we have already computed the coefficients of a factor modulo p^N, then, by Mignotte's theorem, we can obtain the coefficients of a factor in $\mathbb{Q}[x]$.

From now on, we will also assume that p is a prime such that the polynomial $f(x) \pmod p$ is square-free in $\mathbb{F}_p[x]$. Using linear search such a prime p can be found in polynomial time (Corollary 1.77). One can even assume that p is polynomial in the bit size of $f(x)$.

The irreducible factors in $\mathbb{F}_p[x]$ of the polynomial $f(x) \pmod p$ can be found using Berlekamp's deterministic method (Theorem 1.42). Let $g_1(x), \ldots, g_r(x) \in \mathbf{Z}[x]$ be polynomials, all with leading coefficient 1, such that the $g_i(x) \pmod p$ are the irreducible factors of the polynomial $f(x) \pmod p$ in $\mathbb{F}_p[x]$.

Using the technique of Hensel's lemma (Theorem 1.78) and Corollary 1.79, the system $g_1(x), \ldots, g_r(x)$ can be lifted modulo p^N. To simplify the notation, we assume now that $g_1(x), \ldots, g_r(x) \in \mathbf{Z}[x]$ are polynomials with leading coefficients 1 such that

$$f(x) \equiv g_1(x) \cdots g_r(x) \pmod{p^N}$$

and the $g_i(x) \pmod p$ are the irreducible factors of the polynomial $f(x) \pmod p$ in $\mathbb{F}_p[x]$.

Let $h(x) \in \mathbf{Z}[x]$ be an irreducible factor with leading coefficient 1 of the polynomial $f(x)$ in $\mathbb{Q}[x]$. Then there is a uniquely determined set $I \subseteq \{1, \ldots, r\}$ for which

$$h(x) \equiv \prod_{i \in I} g_i(x) \pmod{p^N} .$$

Let N be the smallest integer such that $p^N \geq 2 \cdot 2^{n-1}\|f(x)\|$. Mignotte's bound shows that the polynomial $\prod_{i \in I} g_i(x) \pmod{p^N}$ on the right-hand side, if its coefficients are represented by the residues with the smallest absolute values, coincides with h.

We found that determining the irreducible factors of $f(x)$ is equivalent to finding minimal subsets $I \subseteq \{1, \ldots, r\}$ for which there is a polynomial $h(x) \in \mathbf{Z}[x]$ with leading coefficient 1 such that $h(x) \equiv \prod_{i \in I} g_i(x) \pmod{p^N}$, the absolute values of the coefficients of $h(x)$ are at most $2^{n-1}\|f(x)\|$, and, moreover, $h(x)$ divides $f(x)$. This can be checked by examining at most 2^{r-1} sets I. The cost of examining a single I is polynomial in the size of f.

To summarise, we obtained the following method to factor, in $\mathbb{Q}[x]$, a square-free polynomial $f(x)$ with integer coefficients and leading coefficient 1.

BERLEKAMP-ZASSENHAUS(f)

1 $p \leftarrow$ a prime p such that $f(x) \pmod p$ is square-free in $\mathbb{F}_p[x]$
  and $p = O((n \lg n + 2n \lg \|f\|)^2)$
2 $\{g_1, \ldots, g_r\} \leftarrow$ the irreducible factors of $f(x) \pmod p$ in $\mathbb{F}_p[x]$
  (using Berlekamp's deterministic method)
3 $N \leftarrow \lfloor \log_p(2^{\deg f} \cdot \|f\|) \rfloor + 1$
4 $\{g_1, \ldots, g_r\} \leftarrow$ the Hensel lifting of the system $\{g_1, \ldots, g_r\}$ modulo p^N
5 $\mathcal{I} \leftarrow$ the collection of minimal subsets $I \neq \emptyset$ of $\{1, \ldots r\}$ such that
  $g_I \leftarrow \prod_{i \in I} g_i$ reduced modulo p^N divides f
6 **return** $\{\prod_{i \in I} g_i \ : \ I \in \mathcal{I}\}$

Theorem 1.80 *Let* $f(x) = x^n + \sum_{i=0}^{n-1} a_i x^i \in \mathbf{Z}[x]$ *be a square-free polynomial with integer coefficients and leading coefficient* 1*, and let* p *be a prime number such that the polynomial* $f(x) \pmod p$ *is square-free in* $\mathbb{F}_p[x]$ *and* $p = O((n \lg n + 2n \lg \|f\|)^2)$*. Then the irreducible factors of the polynomial* f *in* $\mathbb{Q}[x]$ *can be obtained by the Berlekamp-Zassenhaus algorithm. The cost of this algorithm is polynomial in* n*,* $\lg \|f(x)\|$ *and* 2^r *where* r *is the number of irreducible factors of the polynomial* $f(x) \pmod p$ *in* $\mathbb{F}_p[x]$*.*

1.5. Example. (Swinnerton-Dyer polynomials) Let

$$f(x) = \prod (x \pm \sqrt{2} \pm \sqrt{3} \pm \cdots \pm \sqrt{p_l}) \in \mathbf{Z}[x] \ ,$$

where $2, 3, \ldots, p_l$ are the first l prime numbers, and the product is taken over all possible 2^l combinations of the signs + and −. The degree of $f(x)$ is $n = 2^l$, and one can show that it is irreducible in $\mathbb{Q}[x]$. On the other hand, for all primes p, the polynomial $f(x) \pmod p$ is the product of factors with degree at most 2. Therefore these polynomials represent hard cases for the Berlekamp-Zassenhaus algorithm, as we need to examine about $2^{n/2-1}$ sets I to find out that f is irreducible.

1.5.3. The LLL algorithm

Our goal in this section is to present the Lenstra-Lenstra-Lovász algorithm (LLL algorithm) for factoring polynomials $f(x) \in \mathbb{Q}[x]$. This was the first polynomial time method for solving the polynomial factorisation problem over $\mathbb{Q}$. Similarly to the Berlekamp-Zassenhaus method, the LLL algorithm starts with a factorisation of f modulo p and then uses Hensel lifting. In the final stages of the work, it uses lattice reduction to find a proper divisor of f, provided one exists. The powerful idea of the LLL algorithm is that it replaced the search, which may have exponential complexity, in the Berlekamp-Zassenhaus algorithm by an efficient lattice reduction.

Let $f(x) \in \mathbf{Z}[x]$ be a square-free polynomial with leading coefficient 1 such that $\deg f = n > 1$, and let p be a prime such that the polynomial $f(x) \pmod p$ is square free in $\mathbb{F}_p[x]$ and $p = O((\lg n + 2n \lg \|f\|)^2)$.

Lemma 1.81 *Suppose that* $f(x) \equiv g_0(x)v(x) \pmod{p^N}$ *where* $g_0(x)$ *and* $v(x)$ *are polynomials with integer coefficients and leading coefficient* 1. *Let* $g(x) \in \mathbf{Z}[x]$ *with* $\deg g(x) = m < n$ *and assume that* $g(x) \equiv g_0(x)u(x) \pmod{p^N}$ *for some polynomial* $u(x)$ *such that* $u(x)$ *has integer coefficients and* $\deg u(x) = \deg g(x) - \deg g_0(x)$. *Let us further assume that* $\|g(x)\|^n \|f(x)\|^m < p^N$. *Then* $\gcd(f(x), g(x)) \neq 1$ *in* $\mathbb{Q}[x]$.

Proof. Let $d = \deg v(x)$. By the assumptions,

$$f(x)u(x) \equiv g_0(x)u(x)v(x) \equiv g(x)v(x) \pmod{p^N} .$$

Suppose that $u(x) = \alpha_0 + \alpha_1 x + \ldots + \alpha_{m-1}x^{m-1}$ and $v(x) = \beta_0 + \beta_1 x + \ldots + \beta_{n-1}x^{n-1}$. (We know that $\beta_d = 1$. If $i > d$, then $\beta_i = 0$, and similarly, if $j > \deg u(x)$, then $\alpha_j = 0$.) Rewriting the congruence, we obtain

$$x^d g(x) + \sum_{j \neq d} \beta_j x^j g(x) - \sum_i \alpha_i x^i f(x) \equiv 0 \pmod{p^N} .$$

Considering the coefficient vectors of the polynomials $x^j g(x)$ and $x^i f(x)$, this congruence amounts to saying that adding to the $(m+d)$-th row of the Sylvester matrix (1.20) a suitable linear combination of the other rows results in a row in which all the elements are divisible by p^N. Consequently, $\det M \equiv 0 \pmod{p^N}$. The Hadamard inequality (Corollary 1.60) yields that $|\det M| \leq \|f\|^m \|g\|^n < p^N$, but this can only happen if $\det M = 0$. However, $\det M = Res(f(x), g(x))$, and so, by (1.21), $\gcd(f(x), g(x)) \neq 1$. ∎

The application of lattice reduction

Set

$$N = \lceil \log_p (2^{2n^2} \|f(x)\|^{2n}) \rceil = O(n^2 + n \lg \|f(x)\|) .$$

Further, we let $g_0(x) \in \mathbf{Z}[x]$ be a polynomial with leading coefficient 1 such that $g_0(x) \pmod{p^N}$ is an irreducible factor of $f(x) \pmod{p^N}$. Set $d = \deg g_0(x) < n$. Define the set L as follows:

$$L = \{g(x) \in \mathbf{Z}[x] : \ \deg g(x) \leq n-1, \ \exists h(x) \in \mathbf{Z}[x], \text{ with } g \equiv hg_0 \pmod{p^N}\} . \quad (1.24)$$

Clearly, L is closed under addition of polynomials. We identify a polynomial with degree less than n with its coefficient vector of length n. Under this identification, L becomes a lattice in $\mathbb{R}^n$. Indeed, it is not too hard to show (Exercise 1.5-2.) that the polynomials

$$p^N 1, p^N x, \ldots, p^N x^{d-1}, \ g_0(x), xg_0(x), \ldots, x^{n-d-1} g_0(x) ,$$

or, more precisely, their coefficient vectors, form a basis of L.

Theorem 1.82 *Let* $g_1(x) \in \mathbf{Z}[x]$ *be a polynomial with degree less than n such that the coefficient vector of* $g_1(x)$ *is the first element in a Lovász-reduced basis of L. Then* $f(x)$ *is irreducible in* $\mathbb{Q}[x]$ *if and only if* $\gcd(f(x), g_1(x)) = 1$.

Proof. As $g_1(x) \neq 0$, it is clear that $\gcd(f(x), g_1(x)) = 1$ whenever $f(x)$ is irreducible. In order to show the implication in the other direction, let us assume that $f(x)$ is reducible and let $g(x)$ be a proper divisor of $f(x)$ such that $g(x) \pmod{p}$ is divisible by $g_0(x) \pmod{p}$ in

$\mathbb{F}_p[x]$. Using Hensel's lemma (Theorem 1.78), we conclude that $g(x) \pmod{p^N}$ is divisible by $g_0(x) \pmod{p^N}$, that is, $g(x) \in L$. Mignotte's theorem (Theorem 1.74) shows that

$$\|g(x)\| \leq 2^{n-1}\|f(x)\| .$$

Now, if we use the properties of reduced bases (second assertion of Theorem 1.67), then we obtain

$$\|g_1(x)\| \leq 2^{(n-1)/2}\|g(x)\| < 2^n\|g(x)\| \leq 2^{2n}\|f(x)\| ,$$

and so

$$\|g_1(x)\|^n\|f(x)\|^{\deg g_1} \leq \|g_1(x)\|^n\|f(x)\|^n < 2^{2n^2}\|f(x)\|^{2n} \leq p^N .$$

We can hence apply Lemma 1.81, which gives $\gcd(g_1(x), f(x)) \neq 1$. ∎

Based on the previous theorem, the LLL algorithm can be outlined as follows (we only give a version for factoring to two factors). The input is a square-free polynomial $f(x) \in \mathbf{Z}[x]$ with integer coefficients and leading coefficient 1 such that $\deg f = n > 1$.

LLL-Polynomial-Factorisation(f)

```
1  p ← a prime p such that f(x) (mod p) is square-free in F_p[x]
       and p = O((n lg n + 2n lg ||f||)^2)
2  w(x) ← an irreducible factor f(x) (mod p) in F_p[x]
       (using Berlekamp's deterministic method)
3  if deg w = n
4     then return "irreducible"
5     else N ← ⌈log_p((2^{2n^2}||f(x)||^{2n})⌉ = O(n^2 + n lg(||f(x)||)
6          (g_0, h_0) ← Hensel-Lifting(f, w, f/w (mod p), p, N)
7          (b_1, ..., b_n) ← a basis of the lattice L ⊆ R^n in (1.24)
8          (g_1, ..., g_n) ← Lovász-Reduction(b_1, ..., b_n)
9          f* ← gcd(f, g_1)
10         if deg f* > 0
11            then return (f*, f/f*)
12            else return "irreducible"
```

Theorem 1.83 *Using the LLL algorithm, the irreducible factors in* $\mathbb{Q}[x]$ *of a polynomial* $f \in \mathbb{Q}[x]$ *can be obtained deterministically in polynomial time.*

Proof. The general factorisation problem, using the method introduced at the discussion of the Berlekamp-Zassenhaus procedure, can be reduced to the case in which the polynomial $f(x) \in \mathbf{Z}[x]$ is square-free and has leading coefficient 1. By the observations made there, the steps in lines 1–7 can be performed in polynomial time. In line 8, the Lovász reduction can be carried out efficiently (Corollary 1.65). In line 9, we may use a modular version of the Euclidean algorithm to avoid intermediate expression swell (see Chapter **??**).

The correctness of the method is asserted by Theorem 1.82. The LLL algorithm can be applied repeatedly to factor the polynomials in the output, in case they are not already irreducible. ∎

One can show that the Hensel lifting costs $O(Nn^2) = O(n^4 + n^3 \lg \|f\|)$ operations with moderately sized integers. The total cost of the version of the LLL algorithm above

is $O(n^5 \lg(p^N)) = O(n^7 + n^6 \lg \|f\|)$.

Exercises

1.5-1 Let $\mathbb{F}$ be a field and let $0 \neq f(x) \in \mathbb{F}[x]$. The polynomial $f(x)$ has no irreducible factors with multiplicity greater than one if and only if $\gcd(f(x), f'(x)) = 1$. *Hint:* In one direction, one can use Lemma 1.13, and use Lemma 1.14 in the other.

1.5-2 Show that the polynomials

$$p^N 1, p^N x, \ldots, p^N x^{d-1}, g_0(x), x g_0(x), \ldots, x^{n-d-1} g_0(x)$$

form a basis of the lattice in (1.24). *Hint:* It suffices to show that the polynomials $p^N x^j$ $(d \leq j < n)$ can be expressed with the given polynomials. To show this, divide $p^N x^j$ by $g_0(x)$ and compute the remainder.

Problems

1-1. The trace in finite fields

Let $\mathbb{F}_{q^k} \supseteq \mathbb{F}_q$ be finite fields. The definition of the trace map $tr = tr_{k,q}$ on $\mathbb{F}_{q^k}$ is as follows: if $\alpha \in \mathbb{F}_{q^k}$ then

$$tr(\alpha) = \alpha + \alpha^q + \cdots + \alpha^{q^{k-1}} .$$

(a) Show that the map tr is $\mathbb{F}_q$-linear and its image is precisely $\mathbb{F}_q$. *Hint:* Use the fact that tr is defined using a polynomial with degree q^{k-1} to show that tr is not identically zero.

(b) Let (α, β) be a uniformly distributed random pair of elements from $\mathbb{F}_{q^k} \times \mathbb{F}_{q^k}$. Then the probability that $tr(\alpha) \neq tr(\beta)$ is $1 - 1/q$.

1-2. The Cantor-Zassenhaus algorithm for fields of characteristic 2

Let $\mathbb{F} = \mathbb{F}_{2^m}$ and let $f(x) \in \mathbb{F}[x]$ be a polynomial of the form

$$f = f_1 f_2 \cdots f_s, \tag{1.25}$$

where the f_i are pairwise relatively prime and irreducible polynomials with degree d in $\mathbb{F}[x]$. Also assume that $s \geq 2$.

(a) Let $u(x) \in \mathbb{F}[x]$ be a uniformly distributed random polynomial with degree less than $\deg f$. Then the greatest common divisor

$$\gcd(u(x) + u^2(x) + \cdots + u^{2^{md-1}}(x), f(x))$$

is a proper divisor of $f(x)$ with probability at least $1/2$.

Hint: Apply the previous exercise taking $q = 2$ and $k = md$, and follow the argument in Theorem 1.38.

(b) Using part (a), give a randomised polynomial time method for factoring a polynomial of the form (1.25) over $\mathbb{F}$.

1-3. Divisors and zero divisors

Let $\mathbb{F}$ be a field. The ring R is said to be an $\mathbb{F}$-***algebra*** (in case $\mathbb{F}$ is clear from the context, R is simply called an algebra), if R is a vector space over $\mathbb{F}$, and $(ar)s = a(rs) = r(as)$ holds for all $r, s \in S$ and $a \in \mathbb{F}$. It is easy to see that the rings $\mathbb{F}[x]$ and $\mathbb{F}[x]/(f)$ are $\mathbb{F}$-algebras.

Let R be a finite-dimensional $\mathbb{F}$-algebra. For an arbitrary $r \in R$, we may consider the map $L_r : R \to R$ defined as $L_r(s) = rs$ for $s \in R$. The map L_r is $\mathbb{F}$-linear, and so we may speak about its minimal polynomial $m_r(x) \in \mathbb{F}[x]$, its characteristic polynomial $k_r(x) \in \mathbb{F}[x]$, and its trace $Tr(r) = Tr(L_r)$. In fact, if U is an ideal in R, then U is an invariant subspace of L_r, and so we can restrict L_r to U, and we may consider the minimal polynomial, the characteristic polynomial, and the trace of the restriction.

(a) Let $f(x),\ g(x) \in \mathbb{F}[x]$ with $\deg\ f > 0$. Show that the residue class $[g(x)]$ is a zero divisor in the ring $\mathbb{F}[x]/(f)$ if and only if f does not divide g and $\gcd(f(x), g(x)) \neq 1$.

(b) Let R be an algebra over $\mathbb{F}$, and let $r \in R$ be an element with minimal polynomial $f(x)$. Show that if f is not irreducible over $\mathbb{F}$, then R contains a zero divisor. To be precise, if $f(x) = g(x)h(x)$ is a non-trivial factorisation ($g, h \in \mathbb{F}[x]$), then $g(r)$ and $h(r)$ form a pair of zero divisors, that is, both of them are non-zero, but their product is zero.

1-4. Factoring polynomials over algebraic number fields

(a) Let $\mathbb{F}$ be a field with characteristic zero and let R be a finite-dimensional $\mathbb{F}$-algebra with an identity element. Let us assume that $R = S_1 \oplus S_2$ where S_1 and S_2 are non-zero $\mathbb{F}$-algebras. Let $r_1, \ldots, r_k$ be a basis of R over $\mathbb{F}$. Show that there is a j such that $m_{r_j}(x)$ is not irreducible in $\mathbb{F}[x]$.

Hint: This exercise is for readers who are familiar with the elements of linear algebra. Let us assume that the minimal polynomial of r_j is the irreducible polynomial $m(x) = x^d - a_1 x^{d-1} + \cdots + a_d$. Let $k_i(x)$ be the characteristic polynomial of L_{r_j} on the invariant subspace U_i (for $i \in \{1, 2\}$). Here U_1 and U_2 are the sets of elements of the form $(s_1, 0)$ and $(0, s_2)$, respectively where $s_i \in S_i$. Because of our conditions, we can find suitable exponents d_i such that $k_i(x) = m(x)^{d_i}$. This implies that the trace $T_i(r_j)$ of the map L_{r_j} on the subspace U_i is $T_i(r_j) = d_i a_1$. Set $e_i = \dim_{\mathbb{F}} U_i$. Obviously, $e_i = d_i d$, which gives $T_1(r_j)/e_1 = T_2(r_j)/e_2$. If the assertion of the exercise is false, then the latter equation holds for all j, and so, as the trace is linear, it holds for all $r \in R$. This, however, leads to a contradiction: if $r = (1, 0) \in S_1 \oplus S_2$ (1 denotes the unity in S_1), then clearly $T_1(r) = e_1$ and $T_2(r) = 0$.

(b) Let $\mathbb{F}$ be an ***algebraic number field***, that is, a field of the form $\mathbb{Q}(\alpha)$ where $\alpha \in \mathbb{C}$, and there is an irreducible polynomial $g(x) \in \mathbf{Z}[x]$ such that $g(\alpha) = 0$. Let $f(x) \in \mathbb{F}[x]$ be a square-free polynomial and set $R = \mathbb{F}[x]/(f)$. Show that R is a finite-dimensional algebra over $\mathbb{Q}$. More precisely, if $\deg\ g = m$ and $\deg\ f = n$, then the elements of the form $\alpha^i [x]^j$ $(0 \leq i < m,\ 0 \leq j < n)$ form a basis over $\mathbb{Q}$.

(c) Show that if f is reducible over $\mathbb{F}$, then there are $\mathbb{Q}$-algebras $S_1,\ S_2$ such that $R \cong S_1 \oplus S_2$.

Hint: Use the Chinese remainder theorem .

(d) Consider the polynomial g above and suppose that a field $\mathbb{F}$ and a polynomial $f \in \mathbb{F}[x]$ are given. Assume, further, that f is square-free and is not irreducible over $\mathbb{F}$. The polynomial f can be factored to the product of two non-constant polynomials in

polynomial time.

Hint: By the previous remarks, the minimal polynomial $m(y)$ over $\mathbb{Q}$ of at least one of the elements $\alpha^i[x]^j$ $(0 \le i \le m,\ 0 \le j \le n)$ is not irreducible in $\mathbb{Q}[y]$. Using the LLL algorithm, $m(y)$ can be factored efficiently in $\mathbb{Q}[y]$. From a factorisation of $m(y)$, a zero divisor of R can be obtained, and this can be used to find a proper divisor of f in $\mathbb{F}[x]$.

Chapter notes

The abstract algebraic concepts discussed in this chapter can be found in many textbooks; see, for instance, Hungerford's book [4].

The theory of finite fields and the related algorithms are the theme of the excellent books by Lidl and Niederreiter [6] and Shparlinski [7].

Our main algorithmic topics, namely the factorisation of polynomials and lattice reduction are thoroughly treated in the book by von zur Gathen and Gerhard [3]. We recommend the same book to the readers who are interested in the efficient methods to solve the basic problems concerning polynomials. Theorem 8.23 of that book estimates the cost of multiplying polynomials by the Schönhage-Strassen method, while Corollary 11.6 is concerned with the cost of the asymptotically fast implementation of the Euclidean algorithm. Ajtai's result about shortest lattice vectors was published in [1].

The method by Kaltofen and Shoup is a randomised algorithm for factoring polynomials over finite fields, and currently it has one of the best time bounds among the known algorithms. The expected number of $\mathbb{F}_q$-operations in this algorithm is $O(n^{1.815} \lg q)$ where $n = \deg f$. Further competitive methods were suggested by von zur Gathen and Shoup, and also by Huang and Pan. The number of operations required by the latter is $O(n^{1.80535} \lg q)$, if $\lg q < n^{0.00173}$. Among the deterministic methods, the one by von zur Gathen and Shoup is the current champion. Its cost is $\widetilde{O}(n^2 + n^{3/2}s + n^{3/2}s^{1/2}p^{1/2})$ operations in $\mathbb{F}_q$ where $q = p^s$. An important related problem is constructing the field $\mathbb{F}_{q^n}$. The fastest randomised method is by Shoup. Its cost is $\tilde{O}(n^2 + n \lg q)$. For finding a square-free factorisation, Yun gave an algorithm that requires $\widetilde{O}(n) + O(n \lg(q/p))$ field operations in $\mathbb{F}_q$.

The best methods to solve the problem of lattice reduction and that of factoring polynomials over the rationals use modular and numerical techniques. After slightly modifying the definition of reduced bases, an algorithm using $\widetilde{O}(n^{3.381} \lg^2 C)$ *bit operations* for the former problem was presented by Storjohann. (We use the original definition introduced in the paper by Lenstra, Lenstra and Lovász [5].) We also mention Schönhage's method using $\widetilde{O}(n^6 + n^4 \lg^2 l)$ bit operations for factoring polynomials with integer coefficients (l is the length of the coefficients).

Besides factoring polynomials with rational coefficients, lattice reduction can also be used to solve lots of other problems: to break knapsack cryptosystems and random number generators based on linear congruences, simultaneous Diophantine approximation, to find integer linear dependencies among real numbers (this problem plays an important rôle in experiments that attempt to find mathematical identities). These and other related problems are discussed in the book [3].

A further exciting application area is the numerical solution of Diophantine equations. One can read about these developments in in the books by Smart [8] and Gaál [2]. The

difficulty of finding a shortest lattice vector was verified in Ajtai's paper [1].

Finally we remark that the practical implementations of the polynomial methods involving lattice reduction are not competitive with the implementations of the Berlekamp-Zassenhaus algorithm, which, in the worst case, has exponential complexity. Nevertheless, the basis reduction performs very well in practice: in fact it is usually much faster than its theoretically proven speed. For some of the problems in the application areas listed above, we do not have another useful method.

The work of the authors was supported in part by grants T042481 and T042706 of the Hungarian Scientific Research Fund.

Bibliography

[1] M. Ajtai. The shortest vector problem in L_2 is NP-hard for randomized reductions. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, 1998, pp. 10–19. 74, 75

[2] I. Gaál. *Diophantine Equations and Power Integral Bases: New Computational Methods*. Birkhäuser Boston, 2002. 74

[3] J. Gathen, von zur, J. Gerhard. *Modern Computer Algebra*. Cambridge University Press, 1999. 74

[4] T. W. Hungerford. *Abstract Algebra: An Introduction*. Saunders College Publishers, 1990. 74

[5] A. K. Lenstra, H. W. Lenstra, Jr., L. Lovász. Factoring polynomials with integer coefficients. *Mathematische Annalen*, 261:513–534, 1982. 74

[6] R. Lidl, H. Niederreiter. *Introduction to Finite Fields and Their Applications*. Cambridge University Press, 1986. 74

[7] I. E. Shparlinski. *Finite Fields: Theory and Computation The Meeting Point of Number Theory, Computer Science, Coding Theory, and Cryptography*. Kluwer Academic Publishers, 1999. 74

[8] N. P. Smart. *The Algorithmic Resolution of Diophantine Equations*. London Mathematical Society Student Text, Vol. 41. Cambridge University Press, 1998. 74

Subject Index

Name index

Contents