JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 16, Number 4, Pages 957–979
S 0894-0347(03)00428-4
Article electronically published on April 25, 2003

SHORT RATIONAL GENERATING FUNCTIONS FOR LATTICE POINT PROBLEMS

ALEXANDER BARVINOK AND KEVIN WOODS

1. Introduction and main results

Our main motivation is the following question, which goes back to Frobenius and Sylvester.

(1.1) The Frobenius Problem. Let $a_1, \ldots, a_d$ be positive coprime integers and let

$$S = \Big\{ \mu_1 a_1 + \cdots + \mu_d a_d : \quad \mu_1, \ldots, \mu_d \in \mathbb{Z}_+ \Big\}$$

be the set of all non-negative integer combinations of $a_1, \ldots, a_d$, or, in other words, the semigroup $S \subset \mathbb{Z}_+$ of non-negative integers generated by $a_1, \ldots, a_d$. What does S look like? In particular, what is the largest integer not in S? (It is well known and easy to see that all sufficiently large integers are in S.) How many positive integers are not in S? How many positive integers within a particular interval or a particular arithmetic progression are not in S?

One of the results of our paper is that for any fixed d "many" of these and similar questions have "easy" solutions. For some of these questions, notably, how to find the largest integer not in S, an efficient solution is already known [K92]. For others, for example, how to find the number of positive integers not in S, an efficient solution was not previously known.

With a subset $S \subset \mathbb{Z}_+$ we associate the *generating function*

$$f(S; x) = \sum_{m \in S} x^m.$$

Clearly, the series converges for all x such that $|x| < 1$. We are interested in finding a "simple" formula for $f(S; x)$.

(1.2) Examples: $d = 2$ and $d = 3$. Suppose that $d = 2$, that is, S is generated by two coprime positive integers a_1 and a_2. It is not hard to show that

$$f(S; x) = \frac{1 - x^{a_1 a_2}}{(1 - x^{a_1})(1 - x^{a_2})}.$$

Suppose that $d = 3$, that is, S is generated by three coprime positive integers a_1, a_2 and a_3. Then there exist (not necessarily distinct) non-negative integers

Received by the editors November 20, 2002.

2000 *Mathematics Subject Classification.* Primary 05A15, 11P21, 13P10, 68W30.

Key words and phrases. Frobenius problem, semigroup, Hilbert series, Hilbert basis, generating functions, computational complexity.

This research was partially supported by NSF Grant DMS 9734138. The second author was partially supported by an NSF VIGRE Fellowship and an NSF Graduate Research Fellowship.

p_1, p_2, p_3, p_4 and p_5, which can be computed efficiently from a_1, a_2 and a_3, such that

$$f(S;x) = \frac{1 - x^{p_1} - x^{p_2} - x^{p_3} + x^{p_4} + x^{p_5}}{(1-x^{a_1})(1-x^{a_2})(1-x^{a_3})}.$$

This interesting fact is, apparently, due to G. Denham [D96]. For example, if $a = 23$, $b = 29$ and $c = 44$, then (thanks to a MAPLE program written by J. Stembridge), $p_1 = 161$, $p_2 = 203$, $p_3 = 220$, $p_4 = 249$ and $p_5 = 335$.

The idea of Denham's proof is to interpret $f(S;x)$ as the Hilbert series of a graded ring $M = \mathbb{C}[t^{a_1}, t^{a_2}, t^{a_3}]$. The Hilbert series of M can be extracted from results of J. Herzog [H70].

We also note that a slightly weaker form of this result is obtained by elementary methods in [SW86].

What happens for $d = 4$ (or larger)? Clearly, since S contains all sufficiently large numbers, $f(S;x)$ is a rational function of the type

$$f(S;x) = p_N(x) + \frac{x^{N+1}}{1-x}, \tag{1.3}$$

where N is the largest integer not in S and $p_N(x)$ is a polynomial of degree N. Can we find a shorter formula for $f(S;x)$?

We need some standard definitions from computational complexity theory (see, for example, [P94]).

(1.4) Definitions. We define the *input size* of an integer a as the number of bits needed to write a, that is, roughly, $1 + \log_2 |a|$. Hence the input size of the sequence $a_1, \ldots, a_d$ will be roughly $d + \sum_{i=1}^{d} \log_2 a_i$. We are interested in the complexity of an algorithm which computes $f(S;x)$ from the input $a_1, \ldots, a_d$. The algorithm is called *polynomial time* provided its running time is bounded by a certain polynomial in the input size.

We show that for any fixed d there is a much shorter formula for $f(S;x)$ than that given by (1.3).

(1.5) Theorem. *Let us fix d. Then there exists a positive integer $s = s(d)$ and a polynomial time algorithm which, given the input $a_1, \ldots, a_d$, computes $f(S;x)$ in the form*

$$f(S;x) = \sum_{i \in I} \alpha_i \frac{x^{p_i}}{(1-x^{b_{i1}}) \cdots (1-x^{b_{is}})},$$

where I is a set of indices, α_i are rational numbers, p_i and b_{ij} are integers and $b_{ij} \neq 0$ for all i, j.

In particular, the number $|I|$ of fractions is bounded by a certain polynomial *poly* in the input size, that is, in $d + \sum_{i=1}^{d} \log_2 a_i$. The degree of *poly* and the number $s = s(d)$ both grow fast with d, roughly as $d^{O(d)}$. However, for any fixed d, the formula of Theorem 1.5 is much shorter than (1.3), in fact, *exponentially shorter*. Indeed, by [EG72] it follows that for any fixed d, the integer N in (1.3) can be as large as $O(t^2)$, where $t = \max\{a_1, \ldots, a_d\}$. Thus the length of formula (1.3) is quadratic in t, that is, *exponential* in the input size. For $d = 4$, there are examples (see [SW86]) showing that if the denominator of $f(S;x)$ is chosen in the form $(1-x^{a_1})(1-x^{a_2})(1-x^{a_3})(1-x^{a_4})$, then the number of monomials in the numerator can grow as fast as $\sqrt{t}$ for $t = \min\{a_1, a_2, a_3, a_4\}$, which is still exponential in the input size.

Theorem 1.5 is a special case of a more general result. Let $S \subset \mathbb{Z}^d$ be a (finite) set of integer points. For an integer vector $m = (\mu_1, \ldots, \mu_d)$ and (complex) variables $\mathbf{x} = (x_1, \ldots, x_d)$, $\mathbf{x} \in \mathbb{C}^d$, let

$$\mathbf{x}^m = x_1^{\mu_1} \cdots x_d^{\mu_d}$$

denote the corresponding monomial. We let $x_i^0 = 1$. Let us consider the Laurent polynomial

$$f(S; \mathbf{x}) = \sum_{m \in S} \mathbf{x}^m.$$

This a priori "long" polynomial can sometimes be written as a "short" rational function

$$f(S; \mathbf{x}) = \sum_{i \in I} \alpha_i \frac{\mathbf{x}^{p_i}}{(1 - \mathbf{x}^{b_{i1}}) \cdots (1 - \mathbf{x}^{b_{ik}})},$$

where $\alpha_i \in \mathbb{Q}$, $p_i, b_{ij} \in \mathbb{Z}^d$ and $b_{ij} \neq 0$ for all i, j. The motivating example is the set $S = \{0, 1, 2, \ldots, n\}$, for which we have

$$f(S; x) = \sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1 - x}.$$

Thus, for this particular S, the long polynomial $f(S; x)$ can be written as a short rational function in x. Indeed, writing $f(S; x)$ as a polynomial requires, roughly, $\Omega(n \log n)$ bits, whereas writing $f(S; x)$ as a rational function requires only $O(\log n)$ bits. A more general example is given by the set of integer points in a rational polyhedron.

(1.6) Definition. Let $c_1, \ldots, c_n \in \mathbb{Z}^d$ be integer vectors and let $\beta_1, \ldots, \beta_n \in \mathbb{Z}$ be integers. The set

$$P = \Big\{ x \in \mathbb{R}^d : \ \langle c_i, x \rangle \leq \beta_i \text{ for } i = 1, \ldots, n \Big\}$$

is called the *rational polyhedron* defined by $\{c_i, \beta_i\}$. Again, we define the input size of P as the number of bits needed to define P. That is, if $c_i = (\gamma_{i1}, \ldots, \gamma_{id})$ then the input size of P is roughly

$$nd + \sum_{i=1}^{n} \log_2 |\beta_i| + \sum_{i=1}^{n} \sum_{j=1}^{d} \log_2 |\gamma_{ij}|.$$

A bounded rational polyhedron is called a *rational polytope.*

In [BP99] it is proved that for any fixed d, if $P \subset \mathbb{R}^d$ is a rational polyhedron which contains no straight lines, then for $S = P \cap \mathbb{Z}^d$ the expression

$$f(S; \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$$

can be written as a short rational function. We give the precise statement in Theorem 3.1.

The main result of this paper is that the *projection* of the set of integer points in a rational polytope has a short generating function as well. More precisely, let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation such that $T(\mathbb{Z}^d) \subset \mathbb{Z}^k$. Thus the matrix of T (which we also denote by T) with respect to the standard bases of $\mathbb{R}^d$ and $\mathbb{R}^k$

is integral. The input size of T is defined similarly as the number of bits needed to write T. Thus, if $T = (t_{ij})$: $i = 1, \ldots, k$ and $j = 1, \ldots, d$, then the input size of T is roughly $kd + \sum_{i=1}^{k} \sum_{j=1}^{d} \log_2 |t_{ij}|$. Let $S = T(P \cap \mathbb{Z}^d)$, $S \subset \mathbb{Z}^k$, be the image of the set of integer points in P. We prove the following result.

(1.7) Theorem. *Let us fix d. There exists a number $s = s(d)$ and a polynomial time algorithm which, given a rational polytope $P \subset \mathbb{R}^d$ and a linear transformation $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ such that $T(\mathbb{Z}^d) \subset \mathbb{Z}^k$, computes the function $f(S; \mathbf{x})$ for $S = T(P \cap \mathbb{Z}^d)$, $S \subset \mathbb{Z}^k$, in the form*

$$f(S; \mathbf{x}) = \sum_{i \in I} \alpha_i \frac{\mathbf{x}^{p_i}}{(1 - \mathbf{x}^{a_{i1}}) \cdots (1 - \mathbf{x}^{a_{is}})},$$

where $\alpha_i \in \mathbb{Q}$, p_i, $a_{ij} \in \mathbb{Z}^k$ and $a_{ij} \neq 0$ for all i, j.

In particular, the number $|I|$ of fractions in the representation of $f(S; \mathbf{x})$ is bounded by a certain polynomial in the input size of P and T. We do not discuss the exact dependence of $s(d)$ on d but note that a rough estimate suggests that s can be chosen about $d^{O(d)}$.

We obtain Theorem 1.5 as a simple corollary of Theorem 1.7 (see Section 6). In Section 7, we discuss other interesting sets which possess short rational generating functions, such as the (minimal) Hilbert bases of rational cones and "test sets" in parametric integer programming. We also discuss a related problem of finding a short formula for the Hilbert series of a ring generated by monomials.

It is not clear at the moment what should be the right version of Theorem 1.7 if we allow P to be an unbounded rational polyhedron: first, it is not clear how to interpret $f(S; \mathbf{x})$ (the defining series may diverge for all $\mathbf{x} \in \mathbb{C}^k$) and second, our methods of Sections 3, 4 and 5 do not work in the case of an infinite S. We note, however, that in various interesting cases (the Frobenius Problem, the Hilbert series of a ring) the case of an unbounded P can be reduced to that of a bounded P, because of a certain "stabilization" in the infinite part of S.

What can we do with rational generating functions? As is discussed in Section 3, we can efficiently perform Boolean operations on sets given by their short rational generating functions. In particular, if $S_1, S_2 \subset \mathbb{Z}^d$ are two finite sets of integer points given by their generating functions $f(S_1; \mathbf{x})$ and $f(S_2; \mathbf{x})$, we can compute the generating functions $f(S_1 \cap S_2; \mathbf{x})$, $f(S_1 \cup S_2; \mathbf{x})$ and $f(S_1 \setminus S_2; \mathbf{x})$ in polynomial time (see Theorem 3.6). Also, by specializing at $\mathbf{x} = (1, \ldots, 1)$, we can count points in polynomial time in finite sets given by their generating functions (this is not immediate since $\mathbf{x} = (1, \ldots, 1)$ is a pole of each fraction in the representation of $f(S; \mathbf{x})$; cf. Theorem 2.6).

Let $f(S; x)$ be the generating function of Theorem 1.5. Then, for the complement $\overline{S} = \mathbb{Z}_+ \setminus S$, we compute the generating function $f(\overline{S}; x) = (1 - x)^{-1} - f(S; x)$ and then compute the number of non-negative integers not in S by specializing $f(\overline{S}; x)$ at $x = 1$. Given an interval $[a, b] \subset \mathbb{Z}_+$, for $S' = S \cap [a, b]$, we can compute $f(S'; x)$, and, specializing at $x = 1$, we can obtain the number of points in S inside the interval $[a, b]$.

The proof of Theorem 1.7 combines several methods. First, it uses some techniques for working with short rational generating functions, developed by the first author; see [BP99] and Sections 2 and 3. Second, it uses some "flatness"-type arguments from the geometry of numbers; see, for example, [GLS93] and Section 4.

Finally, it relies on parametric integer programming arguments developed by R. Kannan, L. Lovász and H. Scarf; see [K92], [KLS90] and Section 5. The crucial step of bringing the three ideas together and obtaining the proof of Theorem 1.7 was taken by the second author (Section 6).

Remark. When a lemma or a theorem states that "there exists a polynomial time algorithm," the actual algorithm is either provided in the proof or a suitable reference is given.

2. Rational functions and monomial substitutions

In this section, we develop certain methods of *specializing* rational functions $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{C}^d$, of the type

$$f(\mathbf{x}) = \sum_{i \in I} \alpha_i \frac{\mathbf{x}^{p_i}}{(1-\mathbf{x}^{a_{i1}}) \cdots (1-\mathbf{x}^{a_{ik(i)}})},$$

where I is a finite set of indices, $\alpha_i \in \mathbb{Q}$, $p_i, a_{ij} \in \mathbb{Z}^d$ and $a_{ij} \neq 0$ for all i, j. We fix an upper bound $k \geq k(i)$ on the number of binomials in every denominator but allow the number of variables d, the number $|I|$ of terms, the coefficients α_i and the vectors p_i, a_{ij} to vary. Moreover, to simplify the notation somewhat, we will consider the case of all $k(i)$ being equal to a number k, so

$$f(\mathbf{x}) = \sum_{i \in I} \alpha_i \frac{\mathbf{x}^{p_i}}{(1-\mathbf{x}^{a_{i1}}) \cdots (1-\mathbf{x}^{a_{ik}})}. \tag{2.1}$$

This is a sufficiently general situation since we can always increase the number of binomials in a fraction by using the identity

$$\begin{aligned}\frac{\mathbf{x}^{p}}{(1-\mathbf{x}^{a_1}) \cdots (1-\mathbf{x}^{a_{k-1}})} &= \frac{\mathbf{x}^{p}(1-\mathbf{x}^{a_k})}{(1-\mathbf{x}^{a_1}) \cdots (1-\mathbf{x}^{a_k})} \\ &= \frac{\mathbf{x}^{p}}{(1-\mathbf{x}^{a_1}) \cdots (1-\mathbf{x}^{a_k})} - \frac{\mathbf{x}^{p+a_k}}{(1-\mathbf{x}^{a_1}) \cdots (1-\mathbf{x}^{a_k})}.\end{aligned}$$

The procedure may increase the number of terms by a factor of 2^k, but since k is assumed to be fixed, this amounts to a constant factor increase.

As usual, the input size of (2.1) is the number of bits needed to write $f(\mathbf{x})$ down.

Let $l_1, \ldots, l_d \in \mathbb{Z}^n$ be integer vectors, $l_i = (\lambda_{i1}, \ldots, \lambda_{in})$. The vectors define the *monomial map* $\phi : \mathbb{C}^n \longrightarrow \mathbb{C}^d$ as follows:

$$\begin{aligned}&\mathbf{z} \longmapsto \mathbf{x} \\ &(z_1, \ldots, z_n) \longmapsto (x_1, \ldots, x_d), \quad \text{where} \quad x_i = \mathbf{z}^{l_i}.\end{aligned} \tag{2.2}$$

The input size of this monomial map is the number of bits needed to define it, that is, roughly, $dn + \sum_{i=1}^{d} \log_2 |\lambda_{ij}|$.

Suppose that the image of ϕ does not consist entirely of poles of $f(\mathbf{x})$. Then we can define a rational function $g : \mathbb{C}^n \longrightarrow \mathbb{C}$ by

$$g(\mathbf{z}) = f\big(\phi(\mathbf{z})\big).$$

The goal of this section is to construct a polynomial time algorithm, which, given a rational function (2.1) with a fixed number k of binomials in each fraction and a monomial substitution (2.2), computes a formula for $g(\mathbf{z})$. Note that we cannot just substitute $\mathbf{x} = \phi(\mathbf{z})$ in the formula (2.1) since any $\mathbf{z} \in \mathbb{C}^n$ may turn out to be a pole for some fraction of (2.1) and yet a regular point of g. For example, if $d = 1$, $n = 0$ and

$$f(x) = \frac{1}{1-x} - \frac{x^{s+1}}{1-x} = \sum_{m=0}^{s} x^m,$$

then $x = 1$ is the pole of both fractions but is a regular point of f; we have $f(1) = s + 1$.

To this end, let us associate with the rational function (2.1) a meromorphic function $F(c)$, $c \in \mathbb{C}^d$, defined by

$$F(c) = \sum_{i \in I} \alpha_i \frac{\exp\langle c, p_i\rangle}{(1 - \exp\langle c, a_{i1}\rangle) \cdots (1 - \exp\langle c, a_{ik}\rangle)}. \tag{2.3}$$

As usual, for $c \in \mathbb{C}^d$ with $c = r + it$, where $r, t \in \mathbb{R}^d$ and $a \in \mathbb{R}^d$, we let $\langle c, a\rangle = \langle r, a\rangle + i\langle t, a\rangle$, where $\langle \cdot, \cdot\rangle$ is the standard scalar product in $\mathbb{R}^d$. The set of poles of the i-th fraction is the union over $1 \leq j \leq k$ of the hyperplanes $\{c \in \mathbb{C}^d : \langle c, a_{ij}\rangle = 0\}$. However, the set of poles of $F(c)$ may be much smaller because of cancellations of singularities.

There is a simple relation between (2.1) and (2.3). For $c = (\gamma_1, \ldots, \gamma_d)$ and $\mathbf{x} = (x_1, \ldots, x_d)$ we write

$$\mathbf{x} = \mathbf{e}^c \quad \text{provided} \quad x_i = \exp\{\gamma_i\} \quad \text{for} \quad i = 1, \ldots, d.$$

Then the functions (2.1) and (2.3) are related by the equation

$$F(c) = f\bigl(\mathbf{e}^c\bigr).$$

Let $L \subset \mathbb{C}^d$ be a subspace such that a generic $c \in L$ is a regular point of $F(c)$. We want to construct a short formula for $F(c)$ for $c \in L$. We assume that the subspace $L \subset \mathbb{C}^d$ is given by its integer basis. Again, we cannot just use (2.3), since L may be orthogonal to some vectors a_{ij} and hence a generic $c \in L$ may be a pole of some fractions in (2.3) while being a regular point of $F(c)$.

(2.4) Definition. Given l, let us consider the function

$$G(\tau; \xi_1, \ldots, \xi_l) = \prod_{i=1}^{l} \frac{\tau \xi_i}{1 - \exp\{-\tau \xi_i\}}$$

in $l + 1$ (complex) variables τ and $\xi_1, \ldots, \xi_l$. It is easy to see that G is analytic in a neighborhood of the origin $\tau = \xi_1 = \cdots = \xi_l = 0$ and therefore there exists an expansion

$$G(\tau; \xi_1, \ldots, \xi_l) = \sum_{j=0}^{+\infty} \tau^j \operatorname{td}_j(\xi_1, \ldots, \xi_l),$$

where $\operatorname{td}_j(\xi_1, \ldots, \xi_l)$ is a homogeneous polynomial of degree j, called the j-th *Todd polynomial* in $\xi_1, \ldots, \xi_l$. It is easy to check that $\operatorname{td}_j(\xi_1, \ldots, \xi_l)$ is a symmetric polynomial with rational coefficients; cf. [BP99].

(2.5) Lemma. *Let us fix k. Then there exists a polynomial time algorithm which, given a function (2.3) and a subspace $L \subset \mathbb{C}^d$ that does not lie entirely in the set of poles of F, computes $F(c)$ for $c \in L$ in the form*

$$F(c) = \sum_{i \in I'} \beta_i \frac{\exp\langle c, q_i\rangle}{\bigl(1 - \exp\langle c, b_{i1}\rangle\bigr) \cdots \bigl(1 - \exp\langle c, b_{is}\rangle\bigr)},$$

where $s \leq k$, $\beta_i \in \mathbb{Q}$, $q_i, b_{ij} \in \mathbb{Z}^d$ and b_{ij} is not orthogonal to L for any i, j.

Proof. Let us consider the representation (2.3). Let us choose a vector $v \in \mathbb{R}^d$ such that $\langle v, a_{ij} \rangle \neq 0$ for all a_{ij}. Such a vector v can be constructed in polynomial time; see, for example, [BP99]. Let τ be a complex parameter. Then, for any regular point c of $F(c)$ the function $F(c + \tau v)$ is an analytic function in a neighborhood of $\tau = 0$ and the constant term of its expansion at $\tau = 0$ is equal to $F(c)$. Hence our goal is to compute the constant term (in τ) of every fraction in the representation (2.3) of $F(c + \tau v)$ and add them up.

Let us consider a typical fraction

$$h(\tau) = \frac{\exp\langle c + \tau v, p \rangle}{\big(1 - \exp\langle c + \tau v, a_1 \rangle\big) \cdots \big(1 - \exp\langle c + \tau v, a_k \rangle\big)},$$

where $p, a_j \in \mathbb{Z}^d$, as a function of τ. Suppose that the vectors a_i orthogonal to L are $a_1, \ldots, a_l$ for some $l \leq k$. Then

$$h(\tau) = \tau^{-l} \exp\langle c, p \rangle \exp\big\{\tau \langle v, p \rangle\big\} \prod_{i=1}^{l} \frac{\tau}{1 - \exp\big\{\tau \langle v, a_i \rangle\big\}} \\ \times \prod_{i=l+1}^{k} \frac{1}{1 - \exp\langle c + \tau v, a_i \rangle}.$$

Now we observe that $\tau^l h(\tau)$ is an analytic function of τ and that our goal is to compute the coefficient of τ^l in the expansion of $\tau^l h(\tau)$ in the neighborhood of $\tau = 0$.

First, we observe that

$$\exp\big\{\tau \langle v, p \rangle\big\} = \sum_{j=0}^{+\infty} \frac{\langle v, p \rangle^j}{j!} \tau^j. \tag{2.5.1}$$

Second, letting $\xi_i = -\langle v, a_i \rangle$ for $i = 1, \ldots, l$, we observe that

$$\prod_{i=1}^{l} \frac{\tau}{1 - \exp\big\{\tau \langle v, a_i \rangle\big\}} = \frac{1}{\xi_1 \cdots \xi_l} \sum_{j=0}^{+\infty} \tau^j \operatorname{td}_j(\xi_1, \ldots, \xi_l). \tag{2.5.2}$$

Finally,

$$\prod_{i=l+1}^{k} \frac{1}{1 - \exp\langle c + \tau v, a_i \rangle} = \sum_{j=0}^{+\infty} H_j(c, a_{l+1}, \ldots, a_k, v) \tau^j \tag{2.5.3}$$

for some functions H_j.

Note that $\tau = 0$ is a regular point of

$$\prod_{i=l+1}^{k} \frac{1}{1 - \exp\langle c + \tau v, a_i \rangle}$$

and so we compute H_j differentiating the product j times and setting $\tau = 0$. By the repeated application of the chain rule, H_j is a polynomial in $\exp\langle c, a_i \rangle$, $\langle v, a_i \rangle$ and $(1 - \exp\langle c, a_i \rangle)^{-1}$. Thus, for all j_1, j_2, j_3 such that $j_1 + j_2 + j_3 = l$, we have to combine the j_1-st term of (2.5.1) , the j_2-nd term of (2.5.2) and the j_3-rd term of (2.5.3). Since $l \leq k$ and k is fixed, we get the desired result. □

Remark. If $L = \{0\}$ and 0 is a regular point of $F(c)$, the algorithm of Lemma 2.5 computes the number $F(0)$. This procedure is used in [B94] to compute the number of integer points in a polytope; see [DH:03] for the practical implementation of the algorithm.

Now we can compute the result of the monomial substitution (2.2) into the rational function (2.1).

(2.6) Theorem. *Let us fix k. Then there exists a polynomial time algorithm which, given a function (2.1) and a monomial map $\phi : \mathbb{C}^n \longrightarrow \mathbb{C}^d$ given by (2.2), such that the image of ϕ does not lie entirely in the set of poles of $f(\mathbf{x})$, computes the function $g(\mathbf{z}) = f\big(\phi(\mathbf{z})\big)$ as*

$$g(\mathbf{z}) = \sum_{i \in I'} \beta_i \frac{\mathbf{z}^{q_i}}{(1-\mathbf{z}^{b_{i1}}) \cdots (1-\mathbf{z}^{b_{is}})},$$

where $s \leq k$, $\beta_i \in \mathbb{Q}$, $q_i, b_{ij} \in \mathbb{Z}^n$ and $b_{ij} \neq 0$ for all i, j.

Proof. Let $F(c)$ be the function (2.3) associated to $f(\mathbf{x})$. With the monomial map (2.2) we associate a linear transformation $\Phi : \mathbb{C}^n \longrightarrow \mathbb{C}^d$

$$c \longmapsto \big(\langle c, l_1 \rangle, \ldots, \langle c, l_d \rangle\big)$$

and the adjoint transformation $\Phi^* : \mathbb{C}^d \longrightarrow \mathbb{C}^n$,

$$\Phi^*(\xi_1, \ldots, \xi_d) = \xi_1 l_1 + \cdots + \xi_d l_d.$$

Let us define

$$G(c) = F\big(\Phi(c)\big) \quad \text{for} \quad c \in \mathbb{C}^n.$$

Hence

$$G(c) = g(\mathbf{e}^c).$$

Let $L \subset \mathbb{C}^d$ be the image of $\mathbb{C}^n$ under Φ. Then L does not lie entirely in the set of poles of $F(c)$. Applying Lemma 2.5, we compute $G(c) = F\big(\Phi(c)\big)$ in the form

$$G(c) = \sum_{i \in I'} \beta_i \frac{\exp\langle \Phi(c), u_i \rangle}{(1 - \exp\langle \Phi(c), v_{i1} \rangle) \cdots (1 - \exp\langle \Phi(c), v_{is} \rangle)},$$

where for i, j we have $\langle \Phi(c), v_{ij} \rangle \neq 0$ for a generic $c \in L$. Now we let $q_i = \Phi^*(u_i)$ and $b_{ij} = \Phi^*(v_{ij})$ so that

$$g(\mathbf{e}^c) = G(c) = \sum_{i \in I'} \beta_i \frac{\exp\langle c, q_i \rangle}{(1 - \exp\langle c, b_{i1} \rangle) \cdots (1 - \exp\langle c, b_{is} \rangle)}$$

and the result follows. □

Remark. In particular, if $\mathbf{x} = (1, \ldots, 1)$ is a regular point of (2.1), we can choose $l_1 = \cdots = l_d = 0$ in (2.2). In this case, the algorithm of Theorem 2.6 computes the value of $f(1, \ldots, 1)$.

3. Operations with generating functions

Some of the results of this section are stated in [BP99]. Many of the proofs in [BP99] are only sketched and some non-trivial details are omitted. We give a mostly independent presentation with complete proofs. The main goal of this section is to prove that if finite sets $S_1, S_2 \subset \mathbb{Z}^d$ are given by their generating functions $f(S_1; \mathbf{x})$ and $f(S_2; \mathbf{x})$, then the generating function $f(S; \mathbf{x})$ of their intersection $S = S_1 \cap S_2$ can be computed efficiently. Our main tool is the generating function for the integer points in a rational polyhedron.

Let $P \subset \mathbb{R}^d$ be a rational polyhedron and let $S = P \cap \mathbb{Z}^d$ be the set of integer points in P. Let

$$f(S; \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m.$$

Thus if P is bounded, $f(S; \mathbf{x})$ is a Laurent polynomial in $\mathbf{x}$. If P (possibly unbounded) does not contain straight lines, then there is a non-empty open set $U \subset \mathbb{C}^d$ such that the series converges absolutely and uniformly on compact subsets of U to a rational function of $\mathbf{x}$. If P contains a straight line, it is convenient to agree that $f(S; \mathbf{x}) \equiv 0$; see [BP99].

We need the following result from [BP99], which states that $f(S; \mathbf{x})$ can be written as a short rational function.

(3.1) Theorem. *Let us fix d. Then there exists a polynomial time algorithm which, for any given rational polyhedron $P \subset \mathbb{R}^d$, computes $f(P \cap \mathbb{Z}^d; \mathbf{x})$ as*

$$f(P \cap \mathbb{Z}^d; \mathbf{x}) = \sum_{i \in I} \epsilon_i \frac{\mathbf{x}^{p_i}}{(1 - \mathbf{x}^{a_{i1}}) \cdots (1 - \mathbf{x}^{a_{id}})},$$

where $\epsilon_i \in \{-1, 1\}$, $p_i, a_{ij} \in \mathbb{Z}^d$, and $a_{ij} \neq 0$ for all i, j. In fact, for each i, $a_{i1}, \ldots, a_{id}$ is a basis of $\mathbb{Z}^d$.

A (complete) proof can be found in [BP99], Theorem 4.4.

To compute the generating function of the intersection of two sets, we compute the result of a more general operation, that is, the Hadamard product of two rational generating functions.

(3.2) Definition. Let g_1 and g_2 be Laurent power series in $\mathbf{x} \in \mathbb{C}^d$,

$$g_1(\mathbf{x}) = \sum_{m \in \mathbb{Z}^d} \beta_{1m} \mathbf{x}^m \quad \text{and} \quad g_2(\mathbf{x}) = \sum_{m \in \mathbb{Z}^d} \beta_{2m} \mathbf{x}^m.$$

The *Hadamard product* $g = g_1 \star g_2$ is the power series

$$g(\mathbf{x}) = \sum_{m \in \mathbb{Z}^d} \beta_m \mathbf{x}^m \quad \text{where} \quad \beta_m = \beta_{1m} \beta_{2m}.$$

First we will show that the Hadamard product of the Laurent expansions of some particular rational functions can be computed in polynomial time. Namely, let us choose a non-zero vector $l \in \mathbb{Z}^d$ and suppose that $a_{11}, \ldots, a_{1k} \in \mathbb{Z}^d$ and $a_{21}, \ldots, a_{2k} \in \mathbb{Z}^d$ are vectors such that $\langle l, a_{ij} \rangle < 0$ for all i, j. Let $p_1, p_2 \in \mathbb{Z}^d$ and let

$$(3.3) \quad g_1(\mathbf{x}) = \frac{\mathbf{x}^{p_1}}{(1 - \mathbf{x}^{a_{11}}) \cdots (1 - \mathbf{x}^{a_{1k}})} \quad \text{and} \quad g_2(\mathbf{x}) = \frac{\mathbf{x}^{p_2}}{(1 - \mathbf{x}^{a_{21}}) \cdots (1 - \mathbf{x}^{a_{2k}})}.$$

We observe that for all $\mathbf{x}$ in a sufficiently small neighborhood U of $\mathbf{x}_0 = \mathbf{e}^l$, we have $|\mathbf{x}^{a_{ij}}| < 1$ and so g_1 and g_2 have Laurent series expansions for $\mathbf{x} \in U$. Indeed, if $|\mathbf{x}^a| < 1$, the fraction $1/(1-\mathbf{x}^a)$ expands as a geometric series

$$\frac{1}{1-\mathbf{x}^a} = \sum_{\mu \in \mathbb{Z}_+} \mathbf{x}^{\mu a},$$

and to obtain the expansions of g_1 and g_2 we multiply the corresponding series. Clearly, the Hadamard product of the expansions converges for all $\mathbf{x} \in U$ to some analytic function h, which we also denote $g_1 \star g_2$. We prove that once the number k of binomials in (3.3) is fixed, there is a polynomial time algorithm for computing the Laurent expansion of $h = g_1 \star g_2$ as a short rational function.

(3.4) Lemma. *Let us fix k. Then there exists a polynomial time algorithm which, given functions (3.3) such that for some $l \in \mathbb{Z}^d$ we have $\langle a_{ij}, l\rangle < 0$ for all i, j, computes a function $h(\mathbf{x})$ in the form*

$$h(\mathbf{x}) = \sum_{i \in I} \beta_i \frac{\mathbf{x}^{q_i}}{(1-\mathbf{x}^{b_{i1}}) \cdots (1-\mathbf{x}^{b_{is}})}$$

with $q_i, b_{ij} \in \mathbb{Z}^d$, $\beta_i \in \mathbb{Q}$ and $s \leq 2k$ such that h has a Laurent expansion in a neighborhood U of $\mathbf{x}_0 = \mathbf{e}^l$ and $h(\mathbf{x}) = g_1(\mathbf{x}) \star g_2(\mathbf{x})$.

Proof. In the space $\mathbb{R}^{2k} = \big\{(\xi_1, \ldots, \xi_{2k})\big\}$ let P be a rational polyhedron defined by the equations

$$p_1 + \xi_1 a_{11} + \cdots + \xi_k a_{1k} = p_2 + \xi_{k+1} a_{21} + \cdots + \xi_{2k} a_{2k}$$

and the inequalities

$$\xi_i \geq 0 \quad \text{for} \quad i = 1, \ldots, 2k.$$

Let $\mathbf{z} = (z_1, \ldots, z_{2k})$ and let us consider the series

$$f(P \cap \mathbb{Z}^{2k}; \mathbf{z}) = \sum_{m \in P \cap \mathbb{Z}^{2k}} \mathbf{z}^m. \tag{3.4.1}$$

Clearly, the series converges absolutely and uniformly on compact sets as long as $|z_i| < 1$ for $i = 1, \ldots, 2k$. By Theorem 3.1 we compute $f(P \cap \mathbb{Z}^{2k}; \mathbf{z})$ in the form

$$f(P \cap \mathbb{Z}^{2k}; \mathbf{z}) = \sum_{i \in I'} \epsilon_i \frac{\mathbf{z}^{u_i}}{(1-\mathbf{z}^{v_{i1}}) \cdots (1-\mathbf{z}^{v_{i(2k)}})}, \tag{3.4.2}$$

for some vectors $u_i, v_{ij} \in \mathbb{Z}^{2k}$ and some numbers $\epsilon_i \in \{-1, 1\}$, where $v_{ij} \neq 0$ for all i, j.

On the other hand, expanding $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ as products of geometric series, we obtain

$$g_1(\mathbf{x}) = \mathbf{x}^{p_1} \prod_{i=1}^{k} \sum_{\mu_i \in \mathbb{Z}_+} \mathbf{x}^{\mu_i a_i} = \sum_{(\mu_1, \ldots, \mu_k) \in \mathbb{Z}_+^k} \mathbf{x}^{p_1 + \mu_1 a_{11} + \cdots + \mu_k a_{1k}} \quad \text{and}$$

$$g_2(\mathbf{x}) = \mathbf{x}^{p_2} \prod_{i=1}^{k} \sum_{\nu_i \in \mathbb{Z}_+} \mathbf{x}^{\nu_i a_i} = \sum_{(\nu_1, \ldots, \nu_k) \in \mathbb{Z}_+^k} \mathbf{x}^{p_2 + \nu_1 a_{21} + \cdots + \nu_k a_{2k}}.$$

Since the Hadamard product is bilinear and since

$$\mathbf{x}^{m_1} \star \mathbf{x}^{m_2} = \begin{cases} \mathbf{x}^{m_1} & \text{if } m_1 = m_2, \\ 0 & \text{if } m_1 \neq m_2, \end{cases}$$

we conclude that

$$g_1(\mathbf{x}) \star g_2(\mathbf{x}) = \mathbf{x}^{p_1} \sum_{\substack{(m,n)\in P\cap\mathbb{Z}^{2k} \\ m=(\mu_1,\dots,\mu_k) \\ n=(\nu_1,\dots,\nu_k)}} \mathbf{x}^{\mu_1 a_{11}+\cdots+\mu_k a_{1k}}.$$

Thus $h(\mathbf{x})$ is obtained from the function $\mathbf{x}^{p_1} f(P \cap \mathbb{Z}^{2k}; \mathbf{z})$ (cf. (3.4.1)–(3.4.2)) by the monomial substitution

$$z_1 = \mathbf{x}^{a_{11}}, \dots, z_k = \mathbf{x}^{a_{1k}}, z_{k+1} = 1, \dots, z_{2k} = 1.$$

Now we use Theorem 2.6 to compute the result of the monomial substitution in (3.4.2). □

Now we are ready to prove the main result of this section. Suppose we have two finite sets $S_1, S_2 \subset \mathbb{Z}^d$ and let $f(S_1; \mathbf{x})$ and $f(S_2; \mathbf{x})$ be the corresponding generating functions

$$f(S_1; \mathbf{x}) = \sum_{m\in S_1} \mathbf{x}^m \quad \text{and} \quad f(S_2; \mathbf{x}) = \sum_{m \in S_2} \mathbf{x}^m.$$

Suppose further, that $f(S_1; \mathbf{x})$ and $f(S_2; \mathbf{x})$ can be written as short rational functions

$$\begin{aligned} f(S_1; \mathbf{x}) &= \sum_{i\in I_1} \alpha_i \frac{\mathbf{x}^{p_i}}{(1-\mathbf{x}^{a_{i1}})\cdots(1-\mathbf{x}^{a_{ik}})} \quad \text{and} \\ f(S_2; \mathbf{x}) &= \sum_{i\in I_2} \beta_i \frac{\mathbf{x}^{q_i}}{(1-\mathbf{x}^{b_{i1}})\cdots(1-\mathbf{x}^{b_{ik}})} \end{aligned} \tag{3.5}$$

with $\alpha_i, \beta_i \in \mathbb{Q}$, $p_i, q_i, a_{ij}, b_{ij} \in \mathbb{Z}^d$ and $a_{ij}, b_{ij} \neq 0$. Now let us consider S_1 and S_2 as *defined* by representations (3.5) of $f(S_1; \mathbf{x})$ and $f(S_2; \mathbf{x})$ as rational functions. Let $S = S_1 \cap S_2$. Our goal is to compute the representation of

$$f(S; \mathbf{x}) = \sum_{m\in S} \mathbf{x}^m$$

as a short rational function. Again, we assume the number of k of binomials in each fraction of (3.5) fixed and allow numbers α_i and β_i, vectors p_i, q_i and a_{ij}, b_{ij} and the number of variables d to vary.

(3.6) Theorem. *Let us fix k. Then there exists a polynomial time algorithm which, given $f_1(S_1; \mathbf{x})$ and $f_2(S_2; \mathbf{x})$, where S_1 and S_2 are finite, computes $f(S; \mathbf{x})$ for $S = S_1 \cap S_2$ in the form*

$$f(S; \mathbf{x}) = \sum_{i\in I} \gamma_i \frac{\mathbf{x}^{u_i}}{(1-\mathbf{x}^{v_{i1}})\cdots(1-\mathbf{x}^{v_{is}})},$$

where $s \leq 2k$, $\gamma_i \in \mathbb{Q}$, $u_i, v_{ij} \in \mathbb{Z}^d$ and $v_{ij} \neq 0$ for all i, j.

Proof. Let us choose a vector $l \in \mathbb{Z}^d$ such that $\langle l, a_{ij}\rangle \neq 0$ and $\langle l, b_{ij}\rangle \neq 0$ for all i, j. As we remarked before, such a vector l can be constructed in polynomial time. When $\langle l, a_{ij}\rangle > 0$ or when $\langle l, b_{ij}\rangle > 0$, we apply the identity

$$\frac{\mathbf{x}^p}{1-\mathbf{x}^a} = -\frac{\mathbf{x}^{p-a}}{1-\mathbf{x}^{-a}},$$

to reverse the direction of a_{ij} or b_{ij}, so that we achieve $\langle l, a_{ij}\rangle < 0$ and $\langle l, b_{ij}\rangle < 0$ for all i, j in the representations (3.5). Then we can write

$$f(S_1;\mathbf{x}) = \sum_{i \in I_1} \alpha_i g_{1i}(\mathbf{x}) \quad \text{and} \quad f(S_2;\mathbf{x}) = \sum_{i \in I_2} \beta_i g_{2i}(\mathbf{x})$$

for some functions g_{i1}, g_{2i} of type (3.3). There are Laurent series expansions of $f(S_1;\mathbf{x})$ and $f(S_2;\mathbf{x})$ in a neighborhood U of the point $\mathbf{x}_0 = \mathbf{e}^l$ and

$$f(S;\mathbf{x}) = f(S_1;\mathbf{x}) \star f(S_2;\mathbf{x}) = \sum_{i_1 \in I_1, i_2 \in I_2} \alpha_{i_1}\beta_{i_2} g_{1i}(\mathbf{x}) \star g_{2i}(\mathbf{x}).$$

We use Lemma 3.4 to compute $f(S;\mathbf{x})$. □

Let $S_1, \ldots, S_m \subset \mathbb{Z}^d$ be sets. We say that $S \subset \mathbb{Z}^d$ is a *Boolean combination* of $S_1, \ldots, S_m$ provided S is obtained from S_i by taking intersections, unions and complements. An immediate corollary of Theorem 3.6 is that the generating function of a Boolean combination of sets can be computed in polynomial time.

(3.7) Corollary. *Let us fix m (the number of sets $S_i \subset \mathbb{Z}^d$) and k (the number of binomials in each fraction of $f(S_i;\mathbf{x})$). Then there exists an $s = s(k, m)$ and a polynomial time algorithm which, for any m finite sets $S_1, \ldots, S_m \subset \mathbb{Z}^d$ given by their generating functions $f(S_i;\mathbf{x})$ and a set $S \subset \mathbb{Z}^d$ defined as a Boolean combination of $S_1, \ldots, S_m$, computes $f(S;\mathbf{x})$ in the form*

$$f(S;\mathbf{x}) = \sum_{i \in I} \gamma_i \frac{\mathbf{x}^{u_i}}{(1-\mathbf{x}^{v_{i1}})\cdots(1-\mathbf{x}^{v_{is}})},$$

where $\gamma_i \in \mathbb{Q}$, $u_i, v_{ij} \in \mathbb{Z}^d$ and $v_{ij} \neq 0$ for all i, j.

Proof. We note that

$$\begin{aligned} &f(S_1 \cup S_2;\mathbf{x}) = f(S_1;\mathbf{x}) + f(S_2;\mathbf{x}) - f(S_1 \cap S_2;\mathbf{x}) \quad \text{and} \\ &f(S_1 \setminus S_2;\mathbf{x}) = f(S_1;\mathbf{x}) - f(S_1 \cap S_2;\mathbf{x}) \end{aligned}$$

for any two subsets $S_1, S_2 \subset \mathbb{Z}^d$. The proof follows by Theorem 3.6. □

Finally, we discuss how to *patch* together several generating functions into a single generating function.

(3.8) Definitions. By the *interior* $\operatorname{int} P$ of a polyhedron $P \subset \mathbb{R}^d$ we always mean the relative interior, that is, the interior of P with respect to its affine hull.

Let $X \subset \mathbb{R}^d$ be a set. We denote by $[X]$ the indicator function $[X] : \mathbb{R}^d \longrightarrow \mathbb{R}$,

$$[X](x) = \begin{cases} 1 & \text{if } x \in X, \\ 0 & \text{if } \ x \notin X. \end{cases}$$

We will need a simple formula for the indicator of the relative interior of a polytope:

$$[\operatorname{int} P] = (-1)^{\dim P} \sum_F (-1)^{\dim F} [F], \tag{3.8.1}$$

where the sum is taken over all faces of P including P itself. This is a simple corollary of the Euler-Poincaré formula; see, for example, Section VI.3 of [B02].

From Theorem 3.1 we deduce the following corollary.

(3.9) Corollary. *Let us fix d. Then there exists a polynomial time algorithm which, for any given rational polytope $P \subset \mathbb{R}^d$, computes $f(S; \mathbf{x})$ with $S = \big(\operatorname{int} P\big) \cap \mathbb{Z}^d$ in the form*

$$f(S; \mathbf{x}) = \sum_{i \in I} \alpha_i \frac{\mathbf{x}^{p_i}}{(1 - \mathbf{x}^{a_{i1}}) \cdots (1 - \mathbf{x}^{a_{id}})},$$

where $\alpha_i \in \mathbb{Q}$, $p_i, a_{ij} \in \mathbb{Z}^d$ and $a_{ij} \neq 0$ for all i, j.

Proof. Applying formula (3.8.1), we get

$$f(S; \mathbf{x}) = (-1)^{\dim P} \sum_F (-1)^{\dim F} f(F \cap \mathbb{Z}^d; \mathbf{x}).$$

Since the dimension d is fixed, there are polynomially many faces F and their descriptions can be computed in polynomial time from the description of P. We use Theorem 3.1 to complete the proof. □

Let us consider the following situation. Let $S \subset \mathbb{Z}^d$ be a finite set and let $Q_1, \ldots, Q_n \subset \mathbb{R}^d$ be a collection of rational polytopes such that $S \subset \bigcup_{i=1}^n \operatorname{int} Q_i$ and $\operatorname{int} Q_i \cap \operatorname{int} Q_j = \emptyset$ for $i \neq j$. In a typical situation, $Q_1, \ldots, Q_n$ is a polytopal complex, that is, the intersection of every two polytopes Q_i and Q_j, if non-empty, is a common face of Q_i and Q_j and a face of a polytope Q_i from the collection is also a polytope from the collection (in particular, not all Q_i are full-dimensional). In this case, $\bigcup_{i=1}^n Q_i = \bigcup_{i=1}^n \operatorname{int} Q_i$, and the $\operatorname{int} Q_i$ are pairwise disjoint.

Suppose that we are given the functions

$$f(S \cap Q_j; \mathbf{x}) = \sum_{i \in I_j} \alpha_{i,j} \frac{\mathbf{x}^{p_{i,j}}}{(1 - \mathbf{x}^{a_{i1,j}}) \cdots (1 - \mathbf{x}^{a_{ik,j}})}$$

and that we want to compute $f(S; \mathbf{x})$. In other words, we want to patch together several generating functions $f(S \cap Q_j; \mathbf{x})$ into a single generating function $f(S; \mathbf{x})$. We obtain the following result.

(3.10) Lemma. *Let us fix k and d. Then there exists a polynomial time algorithm which, given rational polytopes $Q_1, \ldots, Q_n$ with pairwise disjoint interiors and functions $f(S \cap Q_j; \mathbf{x})$ for a finite set $S \subset \bigcup_{i=1}^n \operatorname{int} Q_i$, computes $f(S; \mathbf{x})$ in the form*

$$f(S; \mathbf{x}) = \sum_{i \in I} \beta_i \frac{\mathbf{x}^{q_i}}{(1 - \mathbf{x}^{b_{i1}}) \cdots (1 - \mathbf{x}^{b_{is}})}$$

for $s \leq 2k$.

Proof. We can write

$$f(S; \mathbf{x}) = \sum_{i=1}^n f(S \cap \operatorname{int} Q_i; \mathbf{x}).$$

On the other hand,

$$S \cap \operatorname{int} Q_i = (S \cap Q_i) \cap (\operatorname{int} Q_i \cap \mathbb{Z}^d).$$

First, using Corollary 3.9 we compute $f(\operatorname{int} Q_i \cap \mathbb{Z}^d; \mathbf{x})$, and then using Theorem 3.6 we compute $f(S \cap \operatorname{int} Q_i; \mathbf{x})$. □

4. Lattice width and small gaps

In this section, we establish a simple geometric fact which plays a crucial role in the proof of Theorem 1.7. We start with definitions.

(4.1) Definitions. Let $\Lambda \subset \mathbb{R}^d$ be a lattice (that is, a discrete additive subgroup of $\mathbb{R}^d$ of rank d) and let $\Lambda^* \subset \mathbb{R}^d$ be the dual (reciprocal) lattice, that is,

$$\Lambda^* = \Big\{ c \in \mathbb{R}^d : \ \langle c, x \rangle \in \mathbb{Z} \quad \text{for all} \quad x \in \Lambda \Big\},$$

where $\langle \cdot, \cdot \rangle$ is the standard scalar product in $\mathbb{R}^d$. For a convex body $B \subset \mathbb{R}^d$ (by which we mean a convex compact set) and a non-zero vector $c \in \Lambda^*$ let

$$\operatorname{width}(B, c) = \max_{x \in B} \langle c, x \rangle - \min_{x \in B} \langle c, x \rangle$$

be the *width* of B in the direction of c. Let

$$\operatorname{width}(B) = \inf_{c \in \Lambda^* \setminus \{0\}} \operatorname{width}(B, c)$$

be the *lattice width* of B.

It is known that there exists a constant $\omega(d)$ with the following property: if $B \cap \Lambda = \emptyset$ then $\operatorname{width}(B) \leq \omega(d)$. It is conjectured that $\omega(d) = O(d)$ while the best known value is $\omega(d) = O(d \ln d)$ [BL:99].

We state some obvious properties of the width:

$$\operatorname{width}(B, c) = \operatorname{width}(B + x, c) \quad \text{for any} \quad x \in \mathbb{R}^d \qquad \text{and}$$
$$\operatorname{width}(\alpha B, c) = \alpha \operatorname{width}(B, c) \quad \text{for all} \quad \alpha \geq 0.$$

Consequently,

$$\operatorname{width}(B) = \operatorname{width}(B + x) \quad \text{for any} \quad x \in \mathbb{R}^d \qquad \text{and}$$
$$\operatorname{width}(\alpha B) = \alpha \operatorname{width}(B) \quad \text{for all} \quad \alpha \geq 0.$$

(4.2) Lemma. *Let $B \subset \mathbb{R}^d$ be a convex body, let $c \in \mathbb{R}^d$ be a non-zero vector and let*

$$\gamma_{\min} = \min_{x \in B} \langle c, x \rangle \quad \textit{and} \quad \gamma_{\max} = \max_{x \in B} \langle c, x \rangle.$$

Let $\gamma_{\min} < \gamma_1 < \gamma_2 < \gamma_{\max}$ be numbers. Then there exists a point $x_0 \in B$ and a number $0 < \alpha < 1$ such that for

$$A = \alpha(B - x_0) + x_0 = \alpha B + (1 - \alpha) x_0$$

one has $A \subset B$ and

$$\min_{x \in A} \langle c, x \rangle = \gamma_1 \quad \textit{and} \quad \max_{x \in A} \langle c, x \rangle = \gamma_2.$$

Proof. Translating B, if necessary, we can assume that $\gamma_{\min} = 0$. Dilating B, if necessary, we can assume that $\gamma_{\max} = 1$. Then $0 < \gamma_1/(1 - \gamma_2 + \gamma_1) < 1$, and, therefore, we can choose $x_0 \in B$ such that $\langle c, x_0 \rangle = \gamma_1/(1 - \gamma_2 + \gamma_1)$. Let $\alpha = (\gamma_2 - \gamma_1)$. Then, for $A = \alpha(B - x_0) + x_0 = \alpha B + (1 - \alpha)x_0$, we have

$$\min_{x \in A} \langle c, x \rangle = \frac{(1 - \alpha)\gamma_1}{1 - \gamma_2 + \gamma_1} = \gamma_1$$

and

$$\max_{x \in A} \langle c, x \rangle = \alpha + \frac{(1 - \alpha)\gamma_1}{1 - \gamma_2 + \gamma_1} = \gamma_2.$$

Since B is convex, we have $A \subset B$. □

Now we can prove the main result of this section.

(4.3) Theorem. *Let $B \subset \mathbb{R}^d$ be a convex body and let $\Lambda \subset \mathbb{R}^d$ be a lattice. Let $c \in \Lambda^*$ be a non-zero vector. Consider the map*

$$\phi : B \cap \Lambda \longrightarrow \mathbb{Z}, \qquad \phi(x) = \langle c, x \rangle$$

and let $Y = \phi(B \cap \Lambda)$. Hence $Y \subset \mathbb{Z}$ is a finite set.

Suppose that

$$\operatorname{width}(B, c) \leq 2 \operatorname{width}(B).$$

Then for any $y_1, y_2 \in Y$ such that $y_2 - y_1 > 2\omega(d)$ there exists a $y \in Y$ such that $y_1 < y < y_2$.

Proof. Suppose that such a point y does not exist. Let us choose any $0 < \epsilon < 1/2$ and let $\gamma_1 = y_1 + \epsilon$ and $\gamma_2 = y_2 - \epsilon$. By Lemma 4.2 there exists an $x_0 \in B$ and a number $\alpha > 0$ such that for $A = \alpha(B - x_0) + x_0$, $A \subset B$, we have

$$\min_{x \in A} \langle c, x \rangle = \gamma_1 \quad \text{and} \quad \max_{x \in A} \langle c, x \rangle = \gamma_2.$$

Then there is no integer in the interval $[\gamma_1, \gamma_2]$ which is a value of $\langle c, x \rangle$ for some $x \in B \cap \Lambda$. Hence $A \cap \Lambda = \emptyset$. Therefore, we must have

$$\operatorname{width}(A) \leq \omega(d).$$

On the other hand, since A is a homothetic image of B, we have

$$\operatorname{width}(A) = \alpha \operatorname{width}(B) \quad \text{and} \quad \operatorname{width}(A, c) = \alpha \operatorname{width}(B, c).$$

Therefore,

$$\gamma_2 - \gamma_1 = \operatorname{width}(A, c) \leq 2 \operatorname{width}(A) \leq 2\omega(d).$$

Hence $y_2 - y_1 - 2\epsilon \leq 2\omega(d)$ for any $\epsilon > 0$ and $y_2 - y_1 \leq 2\omega(d)$, which is a contradiction. □

In other words, the set $Y \subset \mathbb{Z}$ does not have "gaps" larger than $2\omega(d)$. We will use the following corollary of Theorem 4.3 (see Section 6.1).

(4.4) Corollary. *Let $Y \subset \mathbb{Z}$ be the set of Theorem 4.3 and let $m = \lceil 2\omega(d) \rceil$. For a positive integer l, let $Y + l = \big\{ y + l : \ y \in Y \big\}$ denote the translation of Y by l. If $Y \neq \emptyset$, then the set*

$$Z = Y \setminus \bigcup_{l=1}^{m} (Y + l)$$

consists of a single point.

Proof. By Theorem 4.3, we have $Z = \{z\}$, where $z = \min\{y : \ y \in Y\}$. □

5. Projections and partitions

In this section, we supply the remaining ingredient of the proof of Theorem 1.7. This ingredient, up to a change of the coordinates, is a weak form of a lemma of R. Kannan [K92].

We describe it below. Let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation such that $T(\mathbb{R}^d) = \mathbb{R}^k$ and $T(\mathbb{Z}^d) \subset \mathbb{Z}^k$. Thus $k \leq d$ and the matrix of T is integral with respect to the standard bases of $\mathbb{R}^d$ and $\mathbb{R}^k$. Then $\ker T$ is a rational $(d-k)$-dimensional subspace of $\mathbb{R}^d$ (that is, a subspace spanned by integer vectors) and $\Lambda = \mathbb{Z}^d \cap (\ker T)$ is a lattice in $\ker T$. As is known (see, for example, Chapter 1 of [C97]), a basis of Λ can be extended to a basis of $\mathbb{Z}^d$ and hence any linear functional

$\ell : \ker T \longrightarrow \mathbb{R}$ such that $\ell(\Lambda) \subset \mathbb{Z}$ can be represented in the form $\ell(x) = \langle c, x\rangle$ for some $c \in \mathbb{Z}^d$. The representation, of course, is not unique as long as $\ker T \neq \mathbb{R}^d$. For $c \in (\ker T)^\perp$ (the orthogonal complement of $\ker T$), the corresponding linear functional is identically 0.

Let $P \subset \mathbb{R}^d$ be a rational polytope. For $y \in \mathbb{R}^k$ let us consider the fiber

$$P_y = \Big\{x \in P : \ T(x) = y\Big\}$$

of x. For $c \in \mathbb{Z}^d \setminus (\ker T)^\perp$ we define the width of P_y in the direction of c as

$$\operatorname{width}(P_y, c) = \max_{x \in P_y} \langle c, x\rangle - \min_{x \in P_y} \langle c, x\rangle$$

and we define the lattice width of P_y as

$$\operatorname{width}(P_y) = \min_{c \in \mathbb{Z}^d \setminus (\ker T)^\perp} \operatorname{width}(P_y, c).$$

We observe that the lattice width of P_y so defined coincides with the width (as defined in Section 4), with respect to Λ, of a translation $P'_y \subset \ker T$.

We need the following result, which is a (rephrased) weaker version of Lemma 3.1 from [K92]. It asserts, roughly, that one can dissect the image $T(P)$ into polynomially many (in the input size of P and T) polyhedral pieces Q_i and find for every piece Q_i a lattice direction w_i such that for all $y \in Q_i$ the lattice width of P_y is almost attained at w_i.

(5.1) Lemma. *Let us fix d. Then there exists a polynomial time algorithm which, for any rational polytope $P \subset \mathbb{R}^d$ and any linear transformation $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ such that $T(\mathbb{R}^d) = \mathbb{R}^k$ and $T(\mathbb{Z}^d) = \mathbb{Z}^k$, constructs rational polytopes $Q_1, \ldots, Q_n \subset \mathbb{R}^k$ and vectors $w_1, \ldots, w_n \in \mathbb{Z}^d \setminus (\ker T)^\perp$ such that*

(1) *For each $i = 1, \ldots, n$ and every $y \in Q_i$,*

$$\textit{either} \quad \operatorname{width}(P_y, w_i) \leq 1 \quad \textit{or} \quad \operatorname{width}(P_y, w_i) \leq 2 \operatorname{width}(P_y);$$

(2) *The interiors $\operatorname{int} Q_i$ are pairwise disjoint and*

$$\bigcup_{i=1}^n \operatorname{int} Q_i = T(P).$$

Proof. Let us construct a rational subspace $V \subset \mathbb{R}^d$ such that $V \cap (\ker T) = \{0\}$ and $(\ker T) + V = \mathbb{R}^d$. Then the restriction of T onto V is invertible and we can compute a matrix L of the linear transformation $\mathbb{R}^k \longrightarrow V$, which is the right inverse of T.

Suppose that the polytope P is defined by a system of linear inequalities

$$P = \Big\{x \in \mathbb{R}^d : \ Ax \leq b\Big\},$$

where A is an $n \times d$ integer matrix and b is an integer n-vector. Then the translation $P'_y \subset \ker T$ of P_y is defined by the system of linear inequalities

$$P'_y = \Big\{x \in \ker T : \ Ax \leq b - ALy\Big\}.$$

As y ranges over $Q = T(P)$, the vector $b' = b - ALy$ ranges over the rational polytope $Q' = b - AL(Q)$ with $\dim Q' \leq k$. Since $\operatorname{width}(P_y, c) = \operatorname{width}(P'_y, c)$ for all $y \in Q$ and all c and $\operatorname{width}(P_y) = \operatorname{width}(P'_y)$, the result follows by Part 3 of Lemma 3.1 of [K92]. □

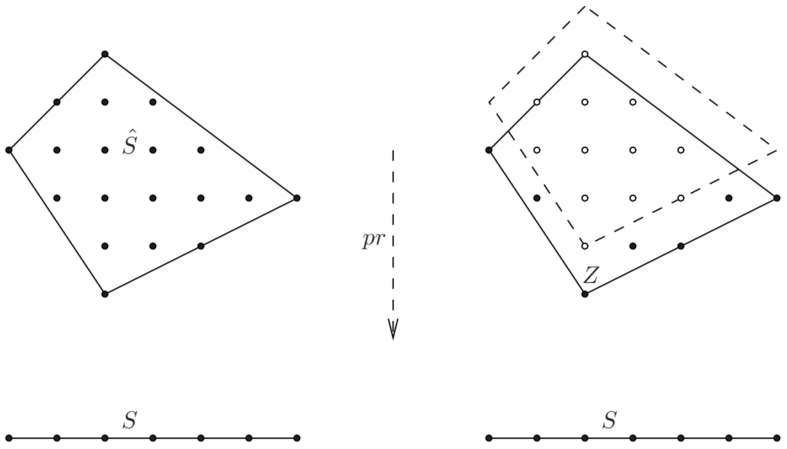

Figure 1.

6. Proofs

Before proving Theorem 1.7, we illustrate one of the main ideas of the proof in the simplest situation.

(6.1) The idea of the proof. Let $pr : \mathbb{R}^{k+1} \longrightarrow \mathbb{R}^k$ be the projection

$$(\xi_1, \ldots, \xi_k, \xi_{k+1}) \longmapsto (\xi_1, \ldots, \xi_k).$$

Let $\hat{S} \subset \mathbb{Z}^{k+1}$ be a finite set, and suppose we know $f\big(\hat{S}; \mathbf{z}\big)$ for $\mathbf{z} = (\mathbf{x}, x_{k+1})$, where $\mathbf{x} \in \mathbb{C}^k$ and $x_{k+1} \in \mathbb{C}$. This situation will occur in the induction step of the proof, and we will want to compute $f(S; \mathbf{x})$, where $S = pr(\hat{S})$. As a special case, suppose that $\hat{S}$ is the set of integer points in a convex polytope, as in Figure 1.

Then the preimage $pr^{-1}(y) \subset \hat{S}$ of every point $y \in S$ is a set of equally spaced points on some interval parallel to the ξ_{k+1}-axis. Let $l = (0, \ldots, 0, 1) \in \mathbb{R}^{k+1}$, let $\hat{S} + l$ be the translation of $\hat{S}$ by l, and let $Z = \hat{S} \setminus \Big(\hat{S} + l\Big)$; see Figure 1. Then the restriction $pr : Z \longrightarrow S$ is necessarily one-to-one and we obtain $f(S; \mathbf{x})$ by specializing $f\big(Z; \mathbf{z}\big)$ at $x_{k+1} = 1$. To compute $f\big(Z; \mathbf{z}\big)$ from $f\big(\hat{S}; \mathbf{z}\big)$ we use Corollary 3.7 and the observation that $f\big(\hat{S} + l; \mathbf{z}\big) = x_{k+1} f\big(\hat{S}; \mathbf{z}\big)$.

In general, the set $\hat{S} \subset \mathbb{R}^{k+1}$ will not be the set of integer points in a polytope, but we will be able to compute $f(\hat{S}, \mathbf{z})$ as a short rational function. The preimage $pr^{-1}(y) \subset \hat{S}$ of a point $y \in S$ will not be a set of equally spaced points, but it will be a set with small gaps; see Section 4. To construct Z we will subtract from $\hat{S}$ not a single, but several translations of $\hat{S}$ to account for all different sizes of gaps.

Proof of Theorem 1.7. Without loss of generality, we assume that $T(\mathbb{R}^d) = \mathbb{R}^k$. Indeed, if $\operatorname{im}(T) \neq \mathbb{R}^k$, we consider the restriction $T : \mathbb{R}^d \longrightarrow \operatorname{im}(T)$. After a change of the coordinates, the lattice $\Lambda = \mathbb{Z}^k \cap \operatorname{im}(T)$ is identified with the standard integer lattice.

The proof is by induction on $\dim(\ker T) = d - k$.

Suppose that $k = d$, so $\dim(\ker T) = 0$ and $T : \mathbb{Z}^d \longrightarrow \mathbb{Z}^k = \mathbb{Z}^d$ is an embedding. Let $e_1, \ldots, e_d$ be the standard basis of $\mathbb{Z}^d$ and let $t_i = T(e_i)$. Then $f(S; \mathbf{x})$ is obtained from $f(P \cap \mathbb{Z}^d; \mathbf{y})$ by the monomial substitution $y_i = \mathbf{x}^{t_i}$ and we use Theorems 2.6 and 3.1 to complete the proof.

Suppose that $d > k$, so $\dim(\ker T) > 0$. Let $Q_1, \ldots, Q_n \subset \mathbb{R}^k$ be the polytopes constructed in Lemma 5.1. It suffices to compute the functions $f(S \cap Q_i; \mathbf{x})$ for $i = 1, \ldots, n$ and then, using Lemma 3.10, we can patch them together and obtain $f(S; \mathbf{x})$.

Let us consider a particular polytope $Q = Q_i$ and the corresponding intersection $S \cap Q$. Let $w = w_i$, $w \in \mathbb{Z}^d \setminus (\ker T)^\perp$ be a vector whose existence is claimed by Lemma 5.1. Let us consider the linear transformation

$$\hat{T} : \mathbb{R}^d \longrightarrow \mathbb{R}^{k+1} = \mathbb{R}^k \oplus \mathbb{R}, \quad \hat{T}(x) = \big(T(x),\ \langle w, x \rangle\big)$$

and the projection

$$pr : \mathbb{R}^{k+1} \longrightarrow \mathbb{R}^k, \quad pr(\xi_1, \ldots, \xi_{k+1}) = (\xi_1, \ldots, \xi_k).$$

Finally, let $P' = \big\{x \in P :\ T(x) \in Q\big\}$ and $\hat{S} = \hat{T}(P' \cap \mathbb{Z}^d) \subset \mathbb{R}^{k+1}$.

Clearly, $S \cap Q = pr(\hat{S})$ and $\dim(\ker \hat{T}) = d - k - 1$, so we can apply the induction hypothesis to $\hat{T}$ and compute $f(\hat{S}; \mathbf{z})$, where $\mathbf{z} = (\mathbf{x}, x_{k+1})$, $x_{k+1} \in \mathbb{C}$. Our goal is to compute $f(S \cap Q; \mathbf{x})$ from $f(\hat{S}; \mathbf{z})$. To do that, we construct a subset $Z \subset \hat{S}$ such that the projection $pr : Z \longrightarrow S \cap Q$ is one-to-one, and then we obtain $f(S \cap Q; \mathbf{x})$ from $f(Z; \mathbf{z})$ by substituting $x_{k+1} = 1$.

For a positive integer l, let $\hat{S} + l$ denote the translation of $\hat{S}$ by l along the last coordinate,

$$\hat{S} + l = \big\{(\xi_1, \ldots, \xi_k, \xi_{k+1} + l) :\ (\xi_1, \ldots, \xi_{k+1}) \in \hat{S}\big\}.$$

Clearly,

$$f(\hat{S} + l; \mathbf{z}) = x_{k+1}^l f(\hat{S}; \mathbf{z}).$$

Let $m = \lceil 2\omega(d-k) \rceil$ (see Section 4) and let us define

$$Z = \hat{S} \setminus \bigcup_{l=1}^{m} (\hat{S} + l).$$

Using Corollary 3.7, we compute $f(Z; \mathbf{z})$.

Now we claim that the projection $pr : Z \longrightarrow S \cap Q$ is one-to-one. Let us consider the projection $pr : \hat{S} \longrightarrow S \cap Q$. For a $y \in S$ let us consider the preimage $\hat{S}_y \subset \hat{S}$ of y. We observe that

$$\hat{S}_y = \Big\{\big(y, \langle w, x \rangle\big) :\ x \in P_y \cap \mathbb{Z}^d\Big\},$$

that is, $\hat{S}_y$ consists of all pairs $\big(y, \langle w, x \rangle\big)$, where x is an integer point from the fiber P_y of P over y:

$$P_y = \Big\{x \in P :\ T(x) = y\Big\}.$$

By Lemma 5.1, we have either $\operatorname{width}(P_y, w) \leq 1$ or $\operatorname{width}(P_y, w) \leq 2 \operatorname{width}(P_y)$. If $\operatorname{width}(P_y, w) \leq 2 \operatorname{width}(P_y)$, then, by Corollary 4.4, the set

$$Z_y = \hat{S}_y \setminus \bigcup_{l=1}^{m} (\hat{S}_y + l)$$

consists of a single point, that is, the point of $\hat{S}_y$ with the smallest last coordinate. If $\operatorname{width}(P_y, w) \leq 1$, then $\hat{S}_y$ consists of a single point and so Z_y consists of a single point as well. Thus, in any case, for any $y \in S \cap Q$ the preimage Z_y of the projection

$pr : Z \longrightarrow S \cap Q$ consists of a single point, so $pr : Z \longrightarrow S \cap Q$ is indeed one-to-one. Hence, using Theorem 2.6, we compute $f(S \cap Q; \mathbf{x})$ by specializing $f(Z; \mathbf{z})$ at $x_{k+1} = 1$ (where $\mathbf{z} = (\mathbf{x}, x_{k+1})$). □

We deduce Theorem 1.5 from Theorem 1.7.

Proof of Theorem 1.5. Let us define a linear transformation $T : \mathbb{R}^d \longrightarrow \mathbb{R}$ by

$$T(\xi_1, \ldots, \xi_d) = a_1\xi_1 + \cdots + a_d\xi_d.$$

Thus $S = T(\mathbb{Z}^d_+)$ is the semigroup generated by $a_1, \ldots, a_d$. It remains to notice that there are some explicit bounds for the largest positive integer not in S, so one can replace the non-negative orthant $\mathbb{Z}^d_+$ by a rational polytope to get the initial interval of S. For example, in [EG72] it is shown that if $t \geq \max\{a_1, \ldots, a_d\}$, then all numbers greater than or equal to $2t^2/d$ are in S. Let $n = \lceil 2t^2/d \rceil$ and let

$$P = \Big\{(\xi_1, \ldots, \xi_d) : \sum_{i=1}^{d} \xi_i a_i \leq n - 1 \text{ and } \xi_i \geq 0 \text{ for } i = 1, \ldots, d\Big\}$$

be the simplex in $\mathbb{R}^d$. Then we can represent S as a disjoint union of $T(P \cap \mathbb{Z}^d)$ and the integer points in the ray $[n, +\infty)$. Since the generating function of the set of integer points in the ray $[n, +\infty)$ is just $x^{n+1}/(1-x)$, applying Theorem 1.7 we complete the proof. □

7. Further examples: Hilbert bases, test sets and Hilbert series

As another application of Theorem 1.7, let us show that certain *Hilbert bases* are enumerated by short rational functions.

Let $u_1, \ldots, u_d \subset \mathbb{Z}^d$ be linearly independent vectors, let

$$\Pi = \Big\{\sum_{i=1}^{d} \alpha_i u_i : \ 0 \leq \alpha_i \leq 1 \text{ for } i = 1, \ldots, d\Big\}$$

be the parallelepiped spanned by $u_1, \ldots, u_d$, and let K be the convex cone spanned by $u_1, \ldots, u_d$:

$$K = \Big\{\sum_{i=1}^{d} \alpha_i u_i : \ \alpha_i \geq 0 \text{ for } i = 1, \ldots, d\Big\}.$$

We say that a point $v \in \Pi \cap \mathbb{Z}^d$ is *indecomposable* provided v cannot be written in the form $v = v_1 + v_2$, where v_1 and v_2 are non-zero integer points from Π. The set S of all indecomposable integer vectors in Π is called the (minimal) *Hilbert basis* of the semigroup $K \cap \mathbb{Z}^d$, since every integer vector in K can be written as a non-negative integer combination of points from S; see Section 16.4 of [Sc86]. Let us show that as long as the dimension d is fixed, the set S has a short rational generating function.

(7.1) Theorem. *Let us fix d. Then there exists a number $s = s(d)$ and a polynomial time algorithm which, given linearly independent vectors $u_1, \ldots, u_d \in \mathbb{Z}^d$, computes the generating function $f(S; \mathbf{x})$ of the (minimal) Hilbert basis S of the semigroup of integer points in the cone spanned by $u_1, \ldots, u_d$ in the form*

$$f(S; x) = \sum_{i \in I} \alpha_i \frac{x^{p_i}}{(1 - x^{b_{i1}}) \cdots (1 - x^{b_{is}})},$$

where I is a set of indices, α_i are rational numbers, p_i, $b_{ij} \in \mathbb{Z}^d$ and $b_{ij} \neq 0$ for all i, j.

Proof. Let us construct a rational polyhedron $Q \subset \Pi$ which contains all integer points in Π except 0. This can be done, for example, as follows: we construct vectors $l_1, \ldots, l_d \in \mathbb{Z}^d$ such that $\langle l_i, u_j \rangle = 0$ for $i \neq j$ and $\langle l_i, u_i \rangle > 0$, let $l = l_1 + \cdots + l_d$ and intersect Π with the half-space $\langle l, x \rangle \geq 1$.

Let $P = Q \times Q \subset \mathbb{R}^d \oplus \mathbb{R}^d = \mathbb{R}^{2d}$ and let $T : P \longrightarrow \mathbb{R}^d$ be the transformation, $T(x, y) = x+y$. Let $S_1 = T(P \cap \mathbb{Z}^{2d})$ and let $S_2 = Q \cap \mathbb{Z}^d$. Then the minimal Hilbert base S can be written as $S = S_2 \setminus S_1$. The proof now follows from Theorem 1.7 and Corollary 3.7. □

Yet another interesting class of sets having short rational generating functions is that of "test sets" with respect to a given integer matrix.

(7.2) Test sets. Let us choose an $n \times d$ integer matrix A such that for any $b \in \mathbb{R}^n$, the polyhedron

$$P_b = \Big\{ x \in \mathbb{R}^d : \ Ax \leq b \Big\},$$

is bounded. A point $h \in \mathbb{Z}^d$, $h \neq 0$, is called a *neighbor* of 0 with respect to A provided there is a polytope P_b containing 0 and h and not containing any other integer point in its interior. The set $S(A)$ of all neighbors of the origin is often called a *test set*. Test sets $S(A)$ play an important role in parametric integer programming [S97]. The set $S(A)$ is finite (assuming that A is sufficiently generic), and it has some interesting (for $d \geq 2$) and not quite understood (for $d \geq 3$) structure. One can show that for any fixed d, given A, the generating function $f(S; \mathbf{x})$ for $S = S(A)$ can be computed in polynomial time as a short rational function. We sketch the argument below.

Let $a_1, \ldots, a_n$ be the rows of A interpreted as vectors from $\mathbb{Z}^d$. If h is a neighbor of 0, then one can choose $b = (\beta_1, \ldots, \beta_n)$ such that P_b contains 0 and h and no integer points in its interior, and each of the inequalities $\langle a_i, x \rangle \leq \beta_i$ is attained as equality either on $x = 0$ or on $x = h$. The hyperplanes $H_i = \Big\{ x \in \mathbb{R}^d : \ \langle a_i, x \rangle = 0 \Big\}$ cut $\mathbb{R}^d$ into polynomially many (in n) polyhedra. Each such polyhedron U is characterized by a subset $I_U \subset \{1, \ldots, n\}$ of the indices i such that $\langle a_i, x \rangle > 0$ for all $x \in \operatorname{int} U$. For a polyhedron U and $h \in U$, let us define $b(h; U) = (\beta_1, \ldots, \beta_n)$, where $\beta_i = \langle a_i, h \rangle$ for $i \in I_U$ and $\beta_i = 0$ for $i \notin I_U$. Thus $b(h; U)$ depends linearly on h. We note that $S \cap U$ is the set of integer points $h \in U$, $h \neq 0$, such that the polytope P_b for $b = b(h; U)$ contains 0 and h and does not contain any other integer point x in its interior. One can see that the set $S \cap U$ can be expressed as a Boolean combination of projections of sets of integer points in some rational polytopes, since the condition $x \in P_b$ for $b = b(h; U)$ can be defined by a system of linear inequalities in x and h. Now the result follows by Lemma 3.10.

We note that other types of test sets studied in the literature, such as Schrijver's universal test set and Graver's test set (see [T95] and [St96]), also admit a short rational generating function.

Finally, we describe one related problem of computational commutative algebra.

(7.3) Hilbert series of rings generated by monomials. Consider integer vectors $a_1, \ldots, a_d \in \mathbb{Z}_+^k$ with non-negative coordinates and let S be the semigroup generated by $a_1, \dots, a_d$:

$$S = \Big\{ \sum_{i=1}^{d} \mu_i a_i : \ \mu_i \in \mathbb{Z}_+ \Big\}.$$

Thus S can be represented as the image $T(\mathbb{Z}_+^d)$ under the linear transformation

$$T : \mathbb{R}^d \longrightarrow \mathbb{R}^k, \quad T(\xi_1, \dots, \xi_d) = \xi_1 a_1 + \cdots + \xi_d a_d.$$

The generating function $f(S; \mathbf{x})$ can be interpreted as the Hilbert series of the $\mathbb{Z}^k$-graded ring $R = \mathbb{C}[\mathbf{x}^{a_1}, \dots, \mathbf{x}^{a_d}]$; cf. [BS98] and Chapter 10 of [St96]. The set S is infinite and Theorem 1.7 is not directly applicable (although it allows us to claim the intersection of S with any given polytopal region has a short rational generating function). However, one can still compute the whole function $f(S; \mathbf{x})$ in polynomial time as a short rational function provided the number d of generators is fixed. We also note that by applying a monomial specialization of $f(S; \mathbf{x})$ we can obtain the Hilbert series of R under a coarser grading.

We sketch an algorithm for computing $f(S; \mathbf{x})$ below.

Without loss of generality we assume that $a_i \neq 0$ for $i = 1, \dots, d$. Consider the product

$$g(S; \mathbf{x}) = f(S; \mathbf{x})(1 - \mathbf{x}^{a_1}) \cdots (1 - \mathbf{x}^{a_d}).$$

It is not hard to prove that $g(S; \mathbf{x})$ is, in fact, a polynomial in $\mathbf{x}$. This follows, for example, from the interpretation of $f(S; \mathbf{x})$ as a Hilbert series; cf. Section I.9 of [E95].

We need to compute a bound L with the property that if the coefficient of $\mathbf{x}^m$, $m = (\mu_1, \dots, \mu_k)$, in $g(S; \mathbf{x})$ is non-zero, then $\mu_1 + \cdots + \mu_k \leq L$. Suppose for a moment that we can find such an L. Let $\mathbb{R}_+^k$ be a non-negative orthant in $\mathbb{R}^k$ and let $\Delta \subset \mathbb{R}_+^k$ be a simplex

$$\Delta = \Big\{ (\mu_1, \dots, \mu_k) \in \mathbb{R}_+^k : \ \mu_1 + \cdots + \mu_k \leq L \Big\}.$$

Then $P = T^{-1}(\Delta) \cap \mathbb{R}_+^d$ is a rational polytope. Let $S' = T(P \cap \mathbb{Z}^d)$, so $S' = \Delta \cap S$. Applying Theorem 1.7, we compute $f(S'; \mathbf{x})$ as a short rational function. Let

$$g(S'; \mathbf{x}) = f(S'; \mathbf{x})(1 - \mathbf{x}^{a_1}) \cdots (1 - \mathbf{x}^{a_d}).$$

We note that

$$g(S; \mathbf{x}) = g(S'; \mathbf{x}) \star f(\Delta; \mathbf{x}).$$

Now we use Theorem 3.1 and Lemma 3.4 to compute the Hadamard product $g(S; \mathbf{x})$ as a short rational function. Finally, we let

$$f(S; \mathbf{x}) = g(S; \mathbf{x}) \prod_{i=1}^{k} \frac{1}{1 - \mathbf{x}^{a_i}}.$$

It remains, therefore, to compute the bound L on the total degree of a monomial $\mathbf{x}^m$ which may appear with a non-zero coefficient in the expansion of $g(S; \mathbf{x})$.

Let us consider the rational cone $K \subset \mathbb{R}^d \oplus \mathbb{R}^d$,

$$K = \Big\{ (x, y) : \ x, y \in \mathbb{R}_+^d \text{ and } T(x) = T(y) \Big\}.$$

The lattice semigroup $K \cap \mathbb{Z}^{2d}$ is finitely generated, and using some standard techniques (see Chapter 17 of [Sc86] and Chapter 4 of [St96]) one can compute in polynomial time an upper bound M on the coordinates of generators (x_i, y_i) of $K \cap \mathbb{Z}^{2d}$.

Let A be the sum of the coordinates of $a_1, \ldots, a_d$. We claim that $L = A(M+1)$ is the desired upper bound.

Indeed, for every generator (x_i, y_i) with $x_i \neq y_i$, let $z_i = x_i - y_i$ or $z_i = y_i - x_i$, whichever is lexicographically positive. Thus each coordinate of z_i is less than or equal to M. Let

$$Z = \mathbb{Z}_+^d \setminus \bigcup_i (\mathbb{Z}_+^d + z_i).$$

One can observe that the restriction $T : Z \longrightarrow S$ is one-to-one. In fact, for every $x \in S$ the vector $z \in Z$ such that $T(z) = x$ is the lexicographic minimum among all $y \in \mathbb{Z}_+^d$ such that $T(y) = x$.

For $I \subset \{1, \ldots, d\}$ let $\mathbb{Z}_+^I \subset \mathbb{Z}_+^d$ be the coordinate semigroup consisting of the points $(\xi_1, \ldots, \xi_d)$ such that $\xi_i = 0$ for $i \notin I$. As is proved in [Kh95], the set Z can be represented as a finite disjoint union of sets Z_j of the type $v_j + \mathbb{Z}_+^{I_j}$ so that the coordinates of v_j do not exceed M. Let $S_j = T(Z_j)$. Then S is the disjoint union of S_j and

$$f(S_j; \mathbf{x}) = \mathbf{x}^{T(v_j)} \prod_{i \in I_j} \frac{1}{1 - \mathbf{x}^{a_i}}.$$

The sum of the coordinates of $T(v_j)$ does not exceed MA. Therefore, if $\mathbf{x}^m$, $m = (\mu_1, \ldots, \mu_k)$, appears with a non-zero coefficient in the product

$$f(S_j; \mathbf{x})(1 - \mathbf{x}^{a_1}) \cdots (1 - \mathbf{x}^{a_k}),$$

we must have $\mu_1 + \cdots + \mu_d \leq MA + A = L$, which completes the proof.

Acknowledgments

The authors are grateful to Ravi Kannan, Herb Scarf and Bernd Sturmfels for many useful discussions and encouragement and to the anonymous referees for their helpful comments.

References

[B94] A. Barvinok, *A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed*, Math. Oper. Res. **19** (1994), 769–779. MR **96c:**52026

[B02] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics, vol. 54, Amer. Math. Soc., Providence, RI, 2002.

[BP99] A. Barvinok and J.E. Pommersheim, *An algorithmic theory of lattice points in polyhedra*, New Perspectives in Algebraic Combinatorics (Berkeley, CA, 1996–97), Math. Sci. Res. Inst. Publ., vol. 38, Cambridge Univ. Press, Cambridge, 1999, pp. 91–147. MR **2000k:**52014

[BS98] D. Bayer and B. Sturmfels, *Cellular resolutions of monomial modules*, J. Reine Angew. Math. **502** (1998), 123–140. MR **99g:**13018

[BL:99] W. Banaszczyk, A.E. Litvak, A. Pajor and S.J. Szarek, *The flatness theorem for nonsymmetric convex bodies via the local theory of Banach spaces*, Math. Oper. Res. **24** (1999), 728–750. MR **2002k:**52019

[C97] J.W.S Cassels, *An Introduction to the Geometry of Numbers. Corrected reprint of the 1971 edition*, Classics in Mathematics, Springer-Verlag, Berlin, 1997. MR **97i:**11074

[D96] G. Denham, *The Hilbert series of a certain module*, manuscript, 1996.

[DH:03] J.A. De Loera, R. Hemmecke, J. Tauzer and R. Yoshida, *Effective lattice point counting in rational convex polytopes*, preprint, http://www.math.ucdavis.edu/∼latte/ (2003).

[E95] D. Eisenbud, *Commutative Algebra with a View Toward Algebraic Geometry*, Graduate Texts in Mathematics, vol. 150, Springer-Verlag, New York, 1995. MR **97a:**13001

[EG72] P. Erdös and R.L. Graham, *On a linear diophantine problem of Frobenius*, Acta Arith. **21** (1972), 399–408. MR **47:**127

[GLS93] M. Grötschel, L. Lovász and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization. Second edition*, Algorithms and Combinatorics, vol. 2, Springer-Verlag, Berlin, 1993. MR **95e:**90001

[H70] J. Herzog, *Generators and relations of abelian semigroups and semigroup rings*, Manuscripta Math. **3** (1970), 175–193. MR **42:**4657

[K92] R. Kannan, *Lattice translates of a polytope and the Frobenius problem*, Combinatorica **12** (1992), 161–177. MR **93k:**52015

[Kh95] A.G. Khovanskii, *Sums of finite sets, orbits of commutative semigroups and Hilbert functions*, Funktsional. Anal. i Prilozhen. **29** (1995), 36–50; English transl. *Funct. Anal. Appl.* **29** (1995), 102–112. MR **96e:**20091

[KLS90] R. Kannan, L. Lovász, and H. Scarf, *The shapes of polyhedra*, Math. Oper. Res. **15** (1990), 364–380. MR **91d:**52004

[P94] C.H. Papadimitriou, *Computational Complexity*, Addison-Wesley, Reading, MA, 1994. MR **95f:**68082

[S97] H. Scarf, *Test sets for integer programs*, Math. Programming, Ser. B **79** (1997), 355–368. MR **98e:**90098

[Sc86] A. Schrijver, *Theory of Linear and Integer Programming*, Wiley-Interscience, Chichester, 1986. MR **88m:**90090

[St96] B. Sturmfels, *Gröbner Bases and Convex Polytopes*, University Lecture Series, vol. 8, Amer. Math. Soc., Providence, RI, 1996. MR **97b:**13034

[SW86] L.A. Székely and N.C. Wormald, *Generating functions for the Frobenius problem with* 2 *and* 3 *generators*, Math. Chronicle **15** (1986), 49–57. MR **88i:**05013

[T95] R. Thomas, *A geometric Buchberger algorithm for integer programming*, Math. Oper. Res. **20** (1995), 864–884. MR **97a:**90027

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109-1109

E-mail address: barvinok@umich.edu

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109-1109

E-mail address: kmwoods@umich.edu