AN ALGORITHMIC THEORY OF LATTICE POINTS IN POLYHEDRA

Alexander Barvinok and James E. Pommersheim

Abstract. The paper discusses various issues related to lattice points in rational polyhedra. The topics include efficient enumeration of lattice points, "short" generating functions for lattice points in rational polyhedra, relations to classical and higher-dimensional Dedekind sums, complexity of the Presburger arithmetic, efficient computations with rational functions, and others. Although the main slant is algorithmic, structural results are discussed, such as relations to the general theory of valuations on polyhedra and connections with the theory of toric varieties. The paper surveys known results and presents some new results and connections.

Contents

Key words and phrases. integer points, lattice, algorithms, polyhedra, toric varieties, generating functions, valuations.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

1. Introduction: "A Formula for The Number of Lattice Points ..."

Let $\mathbb{R}^d$ be Euclidean d-space of all d-tuples $x = (\xi_1, \ldots, \xi_d)$ of real numbers with the standard scalar product $\langle x, y \rangle = \xi_1 \eta_1 + \ldots + \xi_d \eta_d$, where $x = (\xi_1, \ldots, \xi_d)$ and $y = (\eta_1, \ldots, \eta_d)$. The first main object of this paper is the integer lattice $\mathbb{Z}^d \subset \mathbb{R}^d$ consisting of the points with integer coordinates. Let us define the second main object.

(1.1) Definition. A *rational polyhedron* $P \subset \mathbb{R}^d$ is the set of solutions of a finite system of linear inequalities with integer coefficients:

$$P = \Bigl\{ x \in \mathbb{R}^d : \langle c_i, x \rangle \leq \beta_i : \ i = 1, \ldots, m \Bigr\}, \quad \text{where} \quad c_i \in \mathbb{Z}^d \quad \text{and} \quad \beta_i \in \mathbb{Z}.$$

A bounded rational polyhedron is called a *polytope*. A polytope $P \subset \mathbb{R}^d$ is called a *lattice* polytope or an *integer* polytope if its vertices are points from $\mathbb{Z}^d$.

We are interested in the set of lattice points $P \cap \mathbb{Z}^d$ in a given rational polyhedron P. For example, we may be interested in finding a "formula" for the number of lattice points in a given rational or integer polytope P. But what does it mean to "find a formula"? Let us consider a few examples.

(1.2) Example. Suppose that $d = 2$ and $P \subset \mathbb{R}^2$ is an integer polygon. The famous formula of G. Pick [52] states that

$$|P \cap \mathbb{Z}^2| = \operatorname{area}(P) + \frac{|\partial P \cap \mathbb{Z}^d|}{2} + 1,$$

or in words: the number of integer points in an integer polygon is equal to the area of the polygon plus half the number of integer points on the boundary of the polygon plus 1. See [35] as a general reference. We think that nearly every mathematician would agree that Pick's formula is a beautiful and useful formula.

(1.3) Example. Let $P \subset \mathbb{R}^d$ be a polytope. Consider the following formula for the number of lattice points in P:

$$|P \cap \mathbb{Z}^d| = \sum_{x \in \mathbb{Z}^d} \delta(x, P), \qquad \text{where} \quad \delta(x, P) = \begin{cases} 1 & \text{if } x \in P \\ 0 & \text{if } x \notin P. \end{cases}$$

We feel that the majority of mathematicians would agree that this formula is not extremely interesting or useful.

In most cases, the formulae one can get are neither so nice and simple as Pick's formula (Example 1.2), nor so tautological as the formula from Example 1.3. Let us consider a few more examples.

(1.4) Example. Let $\Delta \subset \mathbb{R}^3$ be the tetrahedron with the vertices $(0,0,0)$, $(a,0,0)$, $(0,b,0)$, and $(0,0,c)$, where a, b and c are pairwise coprime positive integers. Then the number of lattice points in Δ can be expressed as

$$|\Delta \cap \mathbb{Z}^3| = \frac{abc}{6} + \frac{ab + ac + bc + a + b + c}{4}$$

$$+ \frac{1}{12}\Bigl(\frac{ac}{b} + \frac{bc}{a} + \frac{ab}{c} + \frac{1}{abc}\Bigr) - s(bc, a) - s(ac, b) - s(ab, c) + 2,$$

where $s(p,q)$ is the *Dedekind sum* defined for coprime positive integers p and q by

$$s(p,q)=\sum_{i=1}^{q}\left(\left(\frac{i}{q}\right)\right)\left(\left(\frac{pi}{q}\right)\right) \quad \text{and} \quad ((x))=\begin{cases} x-\lfloor x\rfloor-\frac{1}{2} & \text{if } x\notin\mathbb{Z} \\ 0 & \text{if } x\in\mathbb{Z}, \end{cases}$$

where $\lfloor\cdot\rfloor$ is the integer part of a number (see [46], [53], [15]).

(1.5) Example. Let $P\subset\mathbb{R}^d$ be a non-empty integer polytope. For a positive integer n, let $nP=\{nx : x\in P\}$ denote the dilated polytope P. As E. Ehrhart discovered (see [20]), there is a polynomial $p(n)$, called the *Ehrhart polynomial* of P, such that

$$|nP\cap\mathbb{Z}^d|=p(n), \quad \text{where} \ \ p(n)=a_dn^d+a_{d-1}n^{d-1}+\ldots+a_0.$$

Furthermore, $a_0=1$, $a_d=\mathrm{vol}_d(P)$, the volume of P, and the following *reciprocity law* holds:

$$p(-n)=(-1)^{\dim(P)}|\mathrm{relint}(nP)\cap\mathbb{Z}^d|, \quad \text{for positive integers } n.$$

That is, the value of p at a negative integer $-n$ equals, up to a sign, the number of integer points in the relative interior of nP. See [61, Section 4.6], for example.

We are going to argue that both Example 1.4 and Example 1.5 are useful and beautiful.

To navigate the sea of "lattice points formulae" which can be found in the literature and which are to be discovered in the future, we have to set up some criteria for beauty and usefulness. Of course, like all such criteria, our criterion is purely subjective. We look at the *computational complexity* of the formula.

Let us fix the space $\mathbb{R}^d$. Suppose that $P\subset\mathbb{R}^d$ is a rational polytope. There is an obvious way to count integer points in P: we consider a sufficiently large box $B=\{x=(\xi_1,\ldots,\xi_d) : \alpha_i\leq\xi_i\leq\beta_i : \ i=1,\ldots,d\}$ which contains P, and check integer points from B one by one to see if they are contained in P. In other words, this is an "effective" version of the formula of Example 1.3. We will measure the "usefulness" and "niceness" of the formula for the number of lattice points by how much time it allows us to save compared with this straightforward procedure of enumeration. In particular, we will be interested in a special class of formulae whose complexity is bounded by a polynomial in the *input size* of the polytope P. A polytope P may be given by its description as a set of linear inequalities (as a rational polyhedron, cf. Definition 1.1). The input size of this *facet* description of P is the total size in binary encoding of the coefficients of the inequalities needed to describe P. For example, the input size of the description $I=\{x : 0\leq x\leq a\}$ of an interval, where a is a positive integer, is $O(\log a)$. A polytope $P\subset\mathbb{R}^d$ may be given as the convex hull of its vertices; the input size of such a *vertex description* is defined in a similar way: the total size of the coordinates of the vertices of P in binary encoding. It is well understood that if the dimension d is fixed and not a part of the input, then the facet description and the vertex description are polynomially equivalent; that is, the length of one is bounded by a polynomial in the length of the other (see, for example, [24], [40], or [58]). Sometimes we talk about formulae that can be applied to polytopes from some particular class. In this case, we are looking at the computational complexity of the formula relative to the input size of the description of P within the class.

From this perspective, we can argue that the formula of Example 1.2 is very nice: it is much more efficient than the direct enumeration of integer points in a polygon. Indeed, it is easy to compute the area of P by triangulating the polygon. Furthermore, the boundary ∂P is a union of finitely many straight line intervals, and counting lattice points in intervals is easy. We would argue that the formula of Example 1.3 is bad since it has exponential complexity even in dimension 1. Indeed, the input size of the interval $[0, a]$ is $O(\log a)$, whereas the straightforward counting of Example 1.3 would give as $O(a)$ complexity, which is exponentially large in the input size. We note that the formula of Example 1.4 is nice, because it reduces counting to computation of the Dedekind sums, which can be computed efficiently. Indeed, by recursively applying the reciprocity relation

$$s(p,q) + s(q,p) = -\frac{1}{4} + \frac{1}{12}\Big(\frac{p}{q} + \frac{q}{p} + \frac{1}{pq}\Big)$$

and the obvious identity

$$s(p,q) = s(r,q), \quad \text{where} \quad r \equiv p (\operatorname{mod} q) \text{ and } \quad 0 \leq r < q$$

(see, for example, [57]), one can compute $s(p,q)$ in time polynomial in the input size of p and q, by a procedure resembling the Euclidean algorithm. Finally, we would argue that the formula of Example 1.5 is also nice since its allows us to save time counting integer points in nP, where n is a large positive integer. Indeed, we can apply the "brute force" counting of Example 1.3 to find the number of integer points in polytopes $P, 2P, \ldots, \lfloor d/2 \rfloor P$ and their relative interiors, and then interpolate the polynomial p. Once p is found, it is easy to find $|nP \cap \mathbb{Z}^d|$ for any positive integer n.

To summarize, we approach every formula in this paper primarily from the computational complexity point of view. Of course, there are different philosophies that are equally legitimate. The topic of this paper is "lattices and polyhedra," as opposed to a close, but somewhat different in spirit, topic "lattices and convex bodies." This is why many interesting results on integer programming, lattice reduction algorithms, and counting lattice points in general convex bodies (see [11], [40], [24], [58]) are not discussed in the paper. Similarly, we do not discuss rather interesting results concerning lattice points in non-rational polyhedra ([59]). Our approach is algebraic and we don't cover recent advances in probabilistic methods of counting (see, for example, [18], [19]). In short, the paper presents "an algorithmic theory," one of many possible.

This area of the research has been quite active. Along with such activity, one expects independent discoveries of certain results, and with unequal publication delays, there is often confusion about who did what first. We have tried to be accurate in the chronology, but, unfortunately, inaccuracies are possible.

This paper is meant to be a survey. However, it does contain some new results: Theorem 4.4 (especially, the second part), Theorem 5.3, Theorem 9.6, results of Section 10, and, possibly, some results of Section 7. In addition, some of the links in Section 8 are new. Whenever possible, we have tried to provide the reader with sketches of proofs.

2. Preliminaries. Algebra of Polyhedra

The number of integer points in a polytope is a *valuation*; that is, it satisfies the inclusion-exclusion property. The theory of valuations, with the theory of the polytope algebra as its basis, was developed by many authors, see [45] and [44] for a survey. Furthermore, several non-equivalent definitions and approaches have been used, each having its own advantages. For example, one can either choose to consider arbitrary polytopes, or to consider lattice polytopes only. In addition, one can decide either to identify or not to identify two polytopes which differ by a (lattice) translation. Also, valuations can be defined via the inclusion-exclusion principle for the union of two or several polytopes. See [30], [32], [36], [43], [47], [48]. Here we employ an approach which is convenient for us.

Let $A \subset \mathbb{R}^d$ be a set. The *indicator function* $[A] : \mathbb{R}^d \longrightarrow \mathbb{R}$ of A is defined as follows:

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

The *algebra of polyhedra* $\mathcal{P}(\mathbb{R}^d)$ is the vector space (over $\mathbb{Q}$) spanned by the indicator functions $[P]$ of all polyhedra $P \subset \mathbb{R}^d$. The space $\mathcal{P}(\mathbb{R}^d)$ is closed under pointwise multiplication of functions: for any two functions $f, g \in \mathcal{P}(\mathbb{R}^d)$, we have $fg \in \mathcal{P}(\mathbb{R}^d)$, since $[P][Q] = [P \cap Q]$. Hence $\mathcal{P}(\mathbb{R}^d)$ is a commutative algebra under pointwise multiplication. We will be interested in some particular subspaces of the algebra of polyhedra. The *polytope algebra* $\mathcal{P}_c(\mathbb{R}^d)$ is the subspace spanned by the indicator functions of all polytopes in $\mathbb{R}^d$, and the *algebra of cones* $\mathcal{P}_K(\mathbb{R}^d)$ is the subspace spanned by the indicator functions of the polyhedral cones in $\mathbb{R}^d$. (A non-empty polyhedron P is called a *cone* provided $\lambda x \in P$ whenever $x \in P$ and $\lambda \geq 0$.)

Clearly, $\mathcal{P}_c(\mathbb{R}^d)$ and $\mathcal{P}_K(\mathbb{R}^d)$ are subalgebras.

(2.1) Definition. A linear transformation

$$\Phi : \mathcal{P}(\mathbb{R}^d) \longrightarrow V,$$

where V is a vector space over $\mathbb{Q}$ is called a *valuation.* Similarly, linear transformations defined on $\mathcal{P}_c(\mathbb{R}^d)$ and $\mathcal{P}_K(\mathbb{R}^d)$ are also called valuations.

One particular valuation is very important.

(2.2) Theorem. *There exists a unique valuation, called the Euler characteristic, $\mu : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathbb{Q}$, such that $\mu([P]) = 1$ for each non-empty polyhedron $P \subset \mathbb{R}^d$.*

Note that we can not simply define μ by letting $\mu([P]) = 1$, because the indicator functions of polyhedra are *not* linearly independent (for $d > 0$). Since the indicator functions $[P]$ span $\mathcal{P}(\mathbb{R}^d)$, the uniqueness is immediate. The following proof belongs to H. Hadwiger. See also [45].

Sketch of Proof. We use induction on d to establish the existence of $\mu = \mu_d$. We have $\mathcal{P}(\mathbb{R}^0) = \mathbb{Q}[0]$, and we can define μ_0 by letting $\mu(a[0]) = a$. Suppose that $d \geq 1$. First, we define μ_d on the polytope subalgebra $\mathcal{P}_c(\mathbb{R}^d)$. Let us choose a non-zero linear function $l : \mathbb{R}^d \longrightarrow \mathbb{R}$ and let us "slice" $\mathbb{R}^d$ into level hyperplanes $H^\alpha = \{x \in \mathbb{R}^d : l(x) = \alpha\}$, with $\alpha \in \mathbb{R}$. For a function $f \in \mathcal{P}_c(\mathbb{R}^d)$, let $f^\alpha : H^\alpha \longrightarrow \mathbb{R}$ be the restriction of f to H^α. We note that H^α can be identified with a $(d-1)$-dimensional Euclidean space, so we can consider the algebra of polytopes $\mathcal{P}_c(H^\alpha)$ and the Euler characteristic $\mu^\alpha : \mathcal{P}_c(H^\alpha) \longrightarrow \mathbb{Q}$. We note that $f^\alpha \in \mathcal{P}_c(H^\alpha)$, and we can define $\mu^\alpha(f^\alpha)$, which we denote by $\mu^\alpha(f)$. For $\alpha \in \mathbb{R}$ and $f \in \mathcal{P}_c(\mathbb{R}^d)$, let us define

$$\mu^{\alpha-}(f) = \lim_{\epsilon \longrightarrow +0} \mu^{\alpha-\epsilon}(f).$$

Suppose that

$$f = \sum_{i \in I} a_i [P_i],$$

where $P_i \subset \mathbb{R}^d$ are polytopes and $a_i \in \mathbb{Q}$ are numbers. It is easy to see that $\mu^{\alpha-}(f)$ is always well-defined, and that $\mu^{\alpha-}(f) = \mu^{\alpha}(f)$ unless α is the minimum value of the linear function l on some P_i. In particular, for any f, there are only finitely many α's for which $\mu^{\alpha-}(f) \neq \mu^{\alpha}(f)$. Now we can define $\mu = \mu_d$ on $\mathcal{P}_c(\mathbb{R}^d)$ by

$$\mu_d(f) = \sum_{\alpha \in \mathbb{R}} \Big(\mu^{\alpha}(f) - \mu^{\alpha-}(f) \Big).$$

The sum is well-defined since there are only finitely many non-zero terms. Now we are ready to extend μ_d to $\mathcal{P}(\mathbb{R}^d)$. Let us choose a polytope Q containing the origin as an interior point, and let $Q(t) = \{tx : x \in Q\}$ be the dilatation of Q by a factor of t. For any $t > 0$ and any $f \in \mathcal{P}(\mathbb{R}^d)$, we have $f_t = [Q(t)]f \in \mathcal{P}_c(\mathbb{R}^d)$ and we let

$$\mu_d(f) = \lim_{t \longrightarrow +\infty} \mu_d(f_t).$$

It is very easy to see that μ_d is well-defined and that it satisfies the condition $\mu_d([P]) = 1$ for any non-empty polyhedron $P \subset \mathbb{R}^d$. □

The Euler characteristic μ allows us to interpret various important valuations as integral transforms with respect to μ *as a measure* (see [34]).

(2.3) Theorem. *Suppose that $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is an affine transformation. Then there is a unique valuation $\mathcal{A} : \mathcal{P}(\mathbb{R}^n) \longrightarrow \mathcal{P}(\mathbb{R}^m)$ such that $\mathcal{A}([P]) = [A(P)]$ for each polyhedron $P \subset \mathbb{R}^d$.*

Proof. Let us define a "kernel" $K : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ by

$$K(x,y) = \begin{cases} 1 & \text{if } \ y = Ax \\ 0 & \text{if } \ y \neq Ax. \end{cases}$$

Then for each fixed y and each $f \in \mathcal{P}(\mathbb{R}^n)$, we have $K(x,y)f \in \mathcal{P}(\mathbb{R}^n)$, so we can apply the Euler characteristic on $\mathcal{P}(\mathbb{R}^n)$, which we denote by μ_x (to stress the variable x). Now we let

$$g = \mathcal{A}(f), \quad \text{where} \quad g(y) = \mu_x\big(K(x,y)f(x)\big).$$

□

In addition to pointwise multiplication, there is a commutative and associative bilinear operation $\star$ on $\mathcal{P}(\mathbb{R}^d)$, which we call *convolution*, because it can be considered as the convolution with respect to the Euler characteristic as a measure. Many authors (cf. [36], [43], [44]) consider $\star$ as the true multiplication in the algebra of polyhedra, and perhaps rightly so, because it has many interesting properties.

(2.4) Definition. Let P and Q be polyhedra in $\mathbb{R}^d$. The *Minkowski sum* $P + Q$ is defined as

$$P + Q = \big\{ x + y : x \in P, \ y \in Q \big\}.$$

(2.5) Theorem. *There is a unique bilinear operation* $\star : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^d)$ *such that* $[P] \star [Q] = [P+Q]$ *for any two polyhedra* P *and* Q.

Proof. Let us fix a decomposition $\mathbb{R}^{2d} = \mathbb{R}^d \times \mathbb{R}^d$. Let $A : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ be the linear transformation $A(x,y) = x+y$ and let $\mathcal{A} : \mathcal{P}(\mathbb{R}^{2d}) \longrightarrow \mathcal{P}(\mathbb{R}^d)$ be the corresponding valuation whose existence is asserted by Theorem 2.3. For functions $f, g \in \mathcal{P}(\mathbb{R}^d)$, let us define their outer product $f \times g \in \mathcal{P}(\mathbb{R}^{2d})$ by $(f \times g)(x,y) = f(x)g(y)$. Then $f \star g = \mathcal{A}(f \times g)$. □

(2.6) Corollary. *Suppose that* $P_1, \ldots, P_k \subset \mathbb{R}^d$ *are polyhedra such that*

$$\alpha_1[P_1] + \ldots + \alpha_k[P_k] = 0$$

for certain rational numbers $\alpha_1, \ldots, \alpha_k$. *Then for any polyhedron* $Q \subset \mathbb{R}^d$, *one has*

$$\alpha_1[P_1 + Q] + \ldots + \alpha_k[P_k + Q] = 0.$$

Proof. We have

$$0 = 0 \star [Q] = \Big(\alpha_1[P_1] + \ldots + \alpha_k[P_k]\Big) \star [Q] = \alpha_1[P_1 + Q] + \ldots + \alpha_k[P_k + Q].$$

□

Convolution $\star$ has many interesting properties (see [45]). For example, it is easy to see that $[0]$ plays the role of the identity. It turns out that $[P]$ is invertible for any polytope P, and that the inverse element is $(-1)^{\dim(P)}[-\text{relint } P]$, the indicator function (up to sign) of the relative interior of the centrally symmetric image $-P$ of P [45].

Next we discuss *duality* in the algebra of cones $\mathcal{P}_K(\mathbb{R}^d)$ (see [36]). If $K \subset \mathbb{R}^d$ is a cone, then

$$K^* = \big\{x \in \mathbb{R}^d : \langle x, y\rangle \geq 0 \quad \text{for each} \quad y \in K\big\}$$

is called the *dual cone* to K.

(2.7) Theorem. *There exists a valuation* $\mathcal{D} : \mathcal{P}_K(\mathbb{R}^d) \longrightarrow \mathcal{P}_K(\mathbb{R}^d)$ *such that*

$$\mathcal{D}([K]) = [\mathbb{R}^d] - [K^*] = [\mathbb{R}^d \setminus K^*]$$

for each cone $K \subset \mathbb{R}^d$.

Sketch of Proof. Let us define the "kernel" $K(x,y) : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ by

$$K(x,y) = \begin{cases} 1 & \text{if } \langle x, y\rangle = -1 \\ 0 & \text{otherwise.} \end{cases}$$

Then for each $f \in \mathcal{P}_K(\mathbb{R}^d)$ and for each y, $K(x,y)f \in \mathcal{P}_K(\mathbb{R}^d)$ and we can apply the Euler characteristic $\mu = \mu_x$. Now we let

$$g = \mathcal{D}(f), \quad \text{where} \quad g(y) = \mu_x\big(K(x,y)f(x)\big).$$

It is straightforward to show that $\mathcal{D}$ satisfies the required properties. □

Theorem 2.7 has an interesting corollary, which says that if a linear identity holds for cones, then the same identity holds for their dual cones.

(2.8) Corollary. *Suppose that $K_1, \ldots, K_m \subset \mathbb{R}^d$ are cones such that*

$$\sum_{i=1}^{m} \alpha_i [K_i] = 0$$

for certain rational numbers α_i. Then

$$\sum_{i=1}^{m} \alpha_i [K_i^*] = 0.$$

Proof. Applying the valuation $\mathcal{D}$ of Theorem 2.7 to both sides of the identity, we get:

$$\sum_{i=1}^{m} \alpha_i [\mathbb{R}^d] - \sum_{i=1}^{m} [K_i^*] = 0.$$

Applying the Euler characteristic μ to both sides of the identity $\sum_{i=1}^{m} \alpha_i [K_i] = 0$, we get that $\sum_{i=1}^{m} \alpha_i = 0$. □

Note, that the valuation $\mathcal{D}$ plays the role of the Fourier Transform with respect to the Euler characteristic μ as a measure. The valuation $\mathcal{D}$ transforms pointwise products into the convolutions:

$$\mathcal{D}(fg) = -\mathcal{D}(f) \star \mathcal{D}(g) \quad \text{for} \quad f, g \in \mathcal{P}_K(\mathbb{R}^d).$$

It suffices to check the identity for $f = [K]$ and $g = [C]$, where $K, C \subset \mathbb{R}^d$ are cones. We have $fg = [K \cap C]$, $\mathcal{D}(f) = [\mathbb{R}^d] - [K^*]$, $\mathcal{D}(g) = [\mathbb{R}^d] - [C^*]$, and

$$\mathcal{D}(f) \star \mathcal{D}(g) = [\mathbb{R}^d] \star [\mathbb{R}^d] - [\mathbb{R}^d] \star [K^*] - [\mathbb{R}^d] \star [C^*] + [K^*] \star [C^*] = -[\mathbb{R}^d] + [K^* + C^*]$$

$$= -[\mathbb{R}^d] + [(K \cap C)^*] = -\mathcal{D}([K \cap C]) = -\mathcal{D}(fg).$$

Finally, we describe an important valuation on the polytope algebra $\mathcal{P}_c(\mathbb{R}^d)$ associated with a vector $u \in \mathbb{R}^d$. Let $P \subset \mathbb{R}^d$ be a polytope and let $u \in \mathbb{R}^d$ be a vector. Let

$$\max(u, P) = \max\big\{\langle u, x \rangle : x \in P\big\},$$

the maximal value of the linear function $\langle u, x \rangle$ on the polytope P and let

$$P_u = \big\{x \in P : \langle u, x \rangle = \max(u, P)\big\}$$

be the face of P where this maximum is attained.

(2.9) Theorem. *For any $u \in \mathbb{R}^d$ there is a valuation*

$$T_u : \mathcal{P}_c(\mathbb{R}^d) \longrightarrow \mathcal{P}_c(\mathbb{R}^d)$$

such that

$$T_u([P]) = [P_u]$$

for any polytope $P \subset \mathbb{R}^d$.

Sketch of Proof. For $\epsilon > 0$ and $\delta > 0$, let us define the kernel

$$K_{\epsilon\delta}(x,y) = \begin{cases} 1 & \text{if } \langle u, x-y \rangle \geq \delta \text{ and } \|x-y\|_\infty \leq \epsilon, \\ 0 & \text{otherwise.} \end{cases}$$

Then for $f \in \mathcal{P}_c(\mathbb{R}^d)$, we let $T_u(f) = g$, where

$$g(y) = f(y) - \lim_{\epsilon \longrightarrow +0} \lim_{\delta \longrightarrow +0} \mu_x \Big(K_{\epsilon\delta}(x,y) f(x) \Big).$$

□

The valuation T_u commutes with convolution: $T_u(f \star g) = T_u(f) \star T_u(g)$. This identity is easy to check on indicator functions of polytopes, and it can then be extended by linearity.

3. Generating Functions for Integer Points in Rational Polyhedra

In this section, we consider the subalgebra $\mathcal{P}(\mathbb{Q}^d) \subset \mathcal{P}(\mathbb{R}^d)$ spanned by the indicator functions $[P]$, where $P \subset \mathbb{Q}^d$ is a *rational* polyhedron. Let $\mathbb{Q}(\mathbf{x})$ be the algebra of rational functions in d complex variables $\mathbf{x} = (x_1, \ldots, x_d)$ with rational coefficients. We discuss a very interesting valuation

$$\mathfrak{F} : \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{Q}(\mathbf{x}).$$

This valuation first was described by J. Lawrence in [37]. At about the same time, A. Khovanskii and A. Pukhlikov gave an independent description in [33]. See also [9].

Let $P \subset \mathbb{R}^d$ be a rational polyhedron. To the set $P \cap \mathbb{Z}^d$ of integral points in P, we associate the generating function

$$f(P \cap \mathbb{Z}^d; \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m, \quad \text{where} \quad \mathbf{x}^m = x_1^{\mu_1} \cdots x_d^{\mu_d} \quad \text{for} \quad m = (\mu_1, \ldots, \mu_d)$$

in d complex variables $\mathbf{x} = (x_1, \ldots, x_d)$. We often write $f(P; \mathbf{x})$ instead of $f(P \cap \mathbb{Z}^d; \mathbf{x})$.

(3.1) Theorem. *There is a map $\mathfrak{F}$ which, to each rational polyhedron $P \subset \mathbb{R}^d$ associates a rational function $f(P; \mathbf{x})$ in d complex variables $\mathbf{x} \in \mathbb{C}^d$, $\mathbf{x} = (x_1, \ldots, x_d)$, such that the following properties are satisfied:*

1. *The map $\mathfrak{F}$ is a valuation: if $P_1, \ldots, P_k \subset \mathbb{R}^d$ are rational polyhedra, such that their indicator functions satisfy a linear identity*
$$\alpha_1[P_1] + \ldots + \alpha_k[P_k] = 0,$$
then the functions $f(P_i; \mathbf{x})$ satisfy the same identity:
$$\alpha_1 f(P_1; \mathbf{x}) + \ldots + \alpha_k f(P_k; \mathbf{x}) = 0.$$
2. *If $m + P$ is a translation of P by an integer vector $m \in \mathbb{Z}^d$, then*
$$f(P + m; \mathbf{x}) = \mathbf{x}^m f(P; \mathbf{x}).$$
3. *We have*
$$f(P; \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$$
for any $\mathbf{x} \in \mathbb{C}^d$ such that the series converges absolutely.
4. *If P contains a straight line then $f(P; \mathbf{x}) \equiv 0$.*

Let us consider some examples.

(3.2) Example. Let us consider the following three rational polyhedra in $\mathbb{R}^1$: $P = \mathbb{R}^1$, $P_+ = \{x : x \geq 0\}$, $P_- = \{x : x \leq 0\}$, and $P_0 = \{0\}$. Part 3 of Theorem 3.1 implies that $f(P_0, x) = x^0 = 1$. Now

$$f(P_+; x) = \sum_{k \geq 0} x^k = \frac{1}{1-x} \quad \text{for} \quad |x| < 1,$$

so by Part 3, we must have $f(P_+; x) = 1/(1-x)$. Similarly,

$$f(P_-; x) = \sum_{k \leq 0} x^k = \frac{1}{1-x^{-1}} = -\frac{x}{1-x} \quad \text{for} \quad |x| > 1,$$

so by Part 3, we must have $f(P_-; x) = -x/(1-x)$. By Part 4, we must have $f(P; x) \equiv 0$. Finally, since $[P] = [P_+] + [P_-] - [P_0]$, Part 1 implies that $f(P_+; x) + f(P_-; x) - f(P_0; x) = 0$, which is indeed the case.

(3.3) Example. Let us choose $k \leq d$ linearly independent integer vectors $u_1, \ldots, u_k$ and let $K = \mathrm{co}\{u_1, \ldots, u_k\}$ be the cone generated by $u_1, \ldots, u_k$. In other words,

$$K = \Big\{\lambda_1 u_1 + \ldots + \lambda_k u_k \quad \text{for some} \quad \lambda_i \geq 0 : \; i = 1, \ldots, k\Big\}.$$

Let

$$\Pi = \Big\{\lambda_1 u_1 + \ldots + \lambda_k u_k \quad \text{for some} \quad 1 > \lambda_i \geq 0 : \; i = 1, \ldots, k\Big\}$$

be the "fundamental parallelepiped" generated by $u_1, \ldots, u_k$. As is well-known (see, for example, [61, Lemma 4.6.7]), for each integer point $m \in K \cap \mathbb{Z}^d$, there is a unique representation

$$m = n + a_1 u_1 + \ldots + a_k u_k,$$

where $n \in \Pi \cap \mathbb{Z}^d$ and $a_1, \ldots, a_k$ are non-negative integers. Let

$$U_K = \Big\{\mathbf{x} \in \mathbb{C}^d : |\mathbf{x}^{u_i}| < 1 \quad \text{for} \quad i = 1, \ldots, k\Big\}.$$

Then $U_K \subset \mathbb{C}^d$ is a non-empty open set, and for each $\mathbf{x} \in U_k$ we have

$$\sum_{m \in K \cap \mathbb{Z}^d} \mathbf{x}^m = \Big(\sum_{n \in \Pi \cap \mathbb{Z}^d} \mathbf{x}^n\Big) \prod_{i=1}^k \frac{1}{1-\mathbf{x}^{u_i}},$$

where the series converges absolutely for each $\mathbf{x} \in U_K$. Part 3 of Theorem 3.1 implies that we must have

$$f(K; \mathbf{x}) = \Big(\sum_{n \in \Pi \cap \mathbb{Z}^d} \mathbf{x}^n\Big) \prod_{i=1}^k \frac{1}{1-\mathbf{x}^{u_i}}.$$

An important particular case arises when the fundamental parallelepiped Π contains only one integer point, the origin. This happens if and only if $u_1, \ldots, u_k$ form a basis of the k-dimensional lattice $\mathrm{Span}\{u_1, \ldots, u_k\} \cap \mathbb{Z}^d$. In this case, the cone K is called *unimodular*, and the function $f(K; \mathbf{x})$ has the especially simple form:

$$f(K; \mathbf{x}) = \prod_{i=1}^{k} \frac{1}{1 - \mathbf{x}^{u_i}}.$$

Theorem 3.1 is very general and powerful, and therefore it has a simple proof. We follow [33], with some changes.

Sketch of Proof of Theorem 3.1. Let us show that if $P \subset \mathbb{R}^d$ is a rational polyhedron without straight lines (or, equivalently, with a vertex) then there exists a non-empty open subset $U_P \subset \mathbb{C}^d$ such that the series

$$\sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$$

converges absolutely for all $\mathbf{x} \in U_P$ to a rational function $f(P; \mathbf{x})$. First, suppose that P is a pointed rational cone, that is

$$P = \big\{x : \langle c_i, x\rangle \leq 0 : \ i = 1, \ldots, m\big\}, \quad \text{where} \quad c_i \in \mathbb{Z}^d$$

and 0 is the vertex of P. Then P can be represented as the conic hull $P = \mathrm{co}\{u_1, \ldots, u_n\}$ of finitely many points $u_i \in \mathbb{Z}^d$, which belong to some open halfspace in $\mathbb{R}^d$. Then

$$U_P = \big\{\mathbf{x} \in \mathbb{C}^d : |\mathbf{x}^{u_i}| < 1 : \ i = 1, \ldots, n\big\}$$

is a non-empty open set, and for each $\mathbf{x} \in U_P$, the series converges absolutely to some function $f(P, \mathbf{x})$. We triangulate P into finitely many simple cones K_i and use the inclusion-exclusion principle to express $f(P; \mathbf{x})$ as a linear combination of $f(K_i; \mathbf{x})$. From Example 3.3, we conclude that $f(P; \mathbf{x})$ is a rational function. Suppose now that $P \subset \mathbb{R}^d$ is an arbitrary rational polyhedron without straight lines. Let us embed $\mathbb{R}^d \subset \mathbb{R}^{d+1}$ by $x \longmapsto (x, 1)$ as the flat $\xi_{d+1} = 1$. Let $K = \mathrm{co}\{P\}$ be the conic hull of P in $\mathbb{R}^{d+1}$. Then K is a pointed rational cone in $\mathbb{R}^{d+1}$, so the function $f\big(K; (\mathbf{x}, t)\big)$: $\mathbf{x} \in \mathbb{C}^d, t \in \mathbb{C}$ is well-defined. Now we observe that

$$f(P; \mathbf{x}) = \frac{\partial f\big(K; (\mathbf{x}, t)\big)}{\partial t}\bigg|_{t=0.}$$

So far, we have defined the map $\mathfrak{F}$ on rational polyhedra $P \subset \mathbb{R}^d$ without straight lines so that Part (3) is satisfied, and so that for every such polyhedron P, the series converges absolutely for all $\mathbf{x}$ in some non-empty open set $U_P \subset \mathbb{C}^d$. Part (2) is then satisfied as well, because it is clearly satisfied for $\mathbf{x} \in U_P$. Finally, if $P_1, \ldots, P_k$ are rational polyhedra without straight lines, then Part (1) is satisfied as long as $U_{P_1} \cap U_{P_2} \cap \ldots \cap U_{P_k} \neq \emptyset$. That is, the functions $f(P_i; \mathbf{x})$ converge for all $\mathbf{x}$ in some non-empty open set in $\mathbb{C}^d$. Now we want to extend $\mathfrak{F}$ to all rational polyhedra P. Let P be an arbitrary rational polyhedron. Let us represent P as a union $P = P_1 \cup P_2 \cup \ldots \cup P_k$, where P_i are polyhedra without straight lines. For $I \subset \{1, \ldots, k\}$, let $P_I = \bigcap_{i \in I} P_i$. Then P_I are rational polyhedra without straight lines. Let us *define* $f(P;\mathbf{x})$ as a linear combination of $f(P_I;\mathbf{x})$ via the inclusion-exclusion principle. Proving that $f(P;\mathbf{x})$ is well-defined boils down to proving that we get a consistent definition of $f(P;\mathbf{x})$ if P itself does not contain straight lines. This is true since there is a non-empty open set $U_P \subset \mathbb{C}^d$, where all the series defining $f(P;\mathbf{x})$ and $f(P_I;\mathbf{x})$ converge absolutely. It then follows that the properties (1)–(3) are satisfied, and it remains to check (4). If P contains a straight line, then for some non-zero $m \in \mathbb{Z}^d$ we have $P + m = P$. Therefore, $f(P;\mathbf{x}) = \mathbf{x}^m f(P;\mathbf{x})$ and $f(P;\mathbf{x})$ must be identically 0. □

The map $\mathfrak{F}$ can be extended to a valuation $\mathfrak{F} : \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{Q}(\mathbf{x})$, which sends every function in $\mathcal{P}_d$ to a rational function in d complex variables, such that

$$\mathfrak{F}(f) = \sum_{m \in \mathbb{Z}^d} f(m)\mathbf{x}^m,$$

provided the series converges absolutely. Furthermore, if $g(x) = f(x - m)$ for some $m \in \mathbb{Z}^d$, then

$$\mathfrak{F}(g) = \mathbf{x}^m \mathfrak{F}(f).$$

Finally, the kernel of this valuation contains the subspace spanned by the indicator functions of rational polyhedra with straight lines.

We are going to present a very interesting and important corollary (Theorem 3.5 below) to Theorem 3.1. This corollary was proved by M. Brion *before* Theorem 3.1 and its first proof used algebraic geometry [6]. Elementary proofs were found by J. Lawrence [37], A.G. Khovanskii and A.V. Pukhlikov [33], A. Barvinok [2] and others. First, we need a definition.

(3.4) Definition. Let $P \subset \mathbb{R}^d$ be a polyhedron and let $v \in P$ be a vertex of P. The *supporting* or *tangent* cone $\operatorname{cone}(P, v)$ of P at v is defined as follows: suppose that

$$P = \big\{x \in \mathbb{R}^d : \langle c_i, x\rangle \leq \beta_i : \ i = 1, \ldots, m\big\}$$

is a representation of P as the set of solutions of a system of linear inequalities, where $c_i \in \mathbb{R}^d$ and $\beta_i \in \mathbb{R}$. Let $I_v = \{i : \langle c_i, v\rangle = \beta_i\}$ be the set of constraints that are active on v. Then

$$\operatorname{cone}(P, v) = \big\{x \in \mathbb{R}^d : \langle c_i, x\rangle \leq \beta_i : \ i \in I_v\big\}.$$

Of course, the cone $\operatorname{cone}(P, v)$ does not depend on a particular system of inequalities chosen to represent P. If P is a rational polyhedron then $\operatorname{cone}(P, v)$ is a rational pointed cone with vertex v. More generally, if $F \subset P$ is a face, we define

$$\operatorname{cone}(P, F) = \big\{x \in \mathbb{R}^d : \langle c_i, x\rangle \leq \beta_i : \ i \in I_F\big\},$$

where I_F is the set of inequalities that are active on F. If $\dim\ P = d$ and $\dim\ F = k$, then the apex of $\operatorname{cone}(P, F)$ is a k-dimensional affine subspace in $\mathbb{R}^d$.

(3.5) Theorem. *Let P be a rational polyhedron. Then*

$$f(P;\mathbf{x}) = \sum_{v \in \operatorname{Vert}(P)} f\big(\operatorname{cone}(P, v);\mathbf{x}\big),$$

where the sum is taken over all vertices v of P.

(3.6) Example. Suppose that $P = \{x : 0 \leq x \leq 1\} \subset \mathbb{R}^1$ is an interval. Then $f(P, x) = x^0 + x^1 = 1 + x$. The polyhedron P has two vertices, 0 and 1, with supporting cones $\operatorname{cone}(P, 0) = [0, +\infty)$ and $\operatorname{cone}(P, 1) = (-\infty, 1]$, respectively. Furthermore,

$$f\big(\operatorname{cone}(P, 0); x\big) = \sum_{m=0}^{+\infty} x^m = \frac{1}{1-x}, \quad \text{and}$$

$$f\big(\operatorname{cone}(P, 1); x\big) = \sum_{m=-\infty}^{1} x^m = \frac{x}{1-x^{-1}} = -\frac{x^2}{1-x}.$$

We observe that indeed

$$f(P; x) = 1 + x = \frac{1}{1-x} - \frac{x^2}{1-x} = f\big(\operatorname{cone}(P, 0); x\big) + f\big(\operatorname{cone}(P, 1); x\big).$$

Sketch of Proof of Theorem 3.5. Let $\mathcal{L} \subset \mathcal{P}(\mathbb{R}^d)$ be the subspace in the polyhedral algebra spanned by the indicator functions $[P]$ of polyhedra that contain straight lines. The theorem follows from Theorem 3.1 and an identity in the polyhedral algebra $\mathcal{P}(\mathbb{R}^d)$: for any polyhedron $P \subset \mathbb{R}^d$

$$[P] \equiv \sum_{v \in \operatorname{Vert}(P)} [\operatorname{cone}(P, v)] \qquad \big(\operatorname{mod} \mathcal{L}\big).$$

First, we demonstrate this identity in the case where $P = \Delta$ is a d-dimensional simplex. Let us represent Δ as the intersection of $d+1$ closed halfspaces $H_1^+, \ldots, H_{d+1}^+$ bounded by flats $H_1, \ldots, H_{d+1}$:

$$\Delta = H_1^+ \cap \ldots \cap H_{d+1}^+.$$

Each tangent cone $\operatorname{cone}(\Delta, v)$ is the intersection of some d halfspaces $H_i^+ : i \in I_v$. Let $H_i^- = \mathbb{R}^d \setminus H_i^+$ be the complementary open halfspaces and let $K_v^- = \bigcap_{i \in I_v} H_i^-$ be the open cone "vertical" to the supporting cone $\operatorname{cone}(\Delta, v)$. Then

$$\mathbb{R}^d \setminus \Delta = \bigcup_{i=1}^{d+1} H_i^-.$$

Let us rewrite $\bigcup_{i=1}^{d+1} H_i^-$ using the inclusion-exclusion formula. Since the intersection of fewer than d halfspaces contains a straight line, we can write

$$[\mathbb{R}^d \setminus \Delta] \equiv (-1)^{d+1} \sum_{v \in \operatorname{Vert}(\Delta)} [K_v^-] \qquad \big(\operatorname{mod} \mathcal{L}\big).$$

Therefore,

$$[\Delta] \equiv (-1)^d \sum_{v \in \operatorname{Vert}(\Delta)} [K_v^-] \qquad \big(\operatorname{mod} \mathcal{L}\big).$$

It remains to show that

$$[\mathrm{cone}(\Delta, v)] \equiv (-1)^d [K_v^-] \qquad \big(\mathrm{mod}\ \mathcal{L}\big).$$

The last assertion follows by comparing $\mathrm{cone}(\Delta, v)$ and K_v^- via a chain of "intermediate" cones $K_v^\epsilon = \bigcap\limits_{i \in I_v} H_i^{\epsilon_i}$, where $\epsilon_i \in \{-,+\}$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_d)$. Note, that $\mathrm{cone}(\Delta, v) = K_v^{+,\ldots,+}$ and $K_v^- = K_v^{-,\ldots,-}$. Once we get

$$[\Delta] \equiv \sum_{v \in \mathrm{Vert}(\Delta)} [\mathrm{cone}(\Delta, v)] \qquad \big(\mathrm{mod}\ \mathcal{L}\big)$$

for any simplex Δ, using triangulations we can show that

$$[P] \equiv \sum_{v \in \mathrm{Vert}(P)} [\mathrm{cone}(P, v)] \qquad \big(\mathrm{mod}\ \mathcal{L}\big)$$

for any polytope P.

Suppose that $P \subset \mathbb{R}^d$ is an arbitrary polyhedron. As is well known, P can be represented as the Minkowski sum $P = Q + K + L$, where Q is a polytope, K is a cone without straight lines, and L is a subspace in $\mathbb{R}^d$. If $L \neq \{0\}$, then P has no vertices, and $f(P; \mathbf{x}) = 0$, so the statement of the theorem is true. If $L = \{0\}$, we may write

$$[Q] \equiv \sum_{v \in \mathrm{Vert}(Q)} [\mathrm{cone}(Q, v)] \qquad \big(\mathrm{mod}\ \mathcal{L}\big),$$

and by Corollary 2.6,

$$[P] \equiv [Q + K] \equiv \sum_{v \in \mathrm{Vert}(Q)} [\mathrm{cone}(Q, v) + K] \qquad \big(\mathrm{mod}\ \mathcal{L}\big).$$

Each vertex v of P is a vertex of Q, and $\mathrm{cone}(P, v) = \mathrm{cone}(Q, v) + K$. Furthermore, a vertex v of Q is a vertex of P if and only if the cone $\mathrm{cone}(Q, v) + K$ does not contain straight lines. The proof now follows. □

Following [9], instead of $\mathfrak{F} : \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{Q}(\mathbf{x})$, one can consider a valuation $\mathfrak{F}' : \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{Q}((\mathbf{x}))$ with the values in the space of formal Laurent power series in $\mathbf{x} = (x_1, \ldots, x_d)$ with rational coefficients:

$$\mathfrak{F}'(f) = \sum_{m \in \mathbb{Z}^d} f(m) \mathbf{x}^m.$$

This leads to essentially the same theory: the space $\mathbb{Q}((\mathbf{x}))$ has the natural structure of a module over the ring of polynomials $\mathbb{Q}[x_1, \ldots, x_d]$. A formal power series f is identified with a rational function $g(\mathbf{x})/h(\mathbf{x})$, where $h(\mathbf{x}) = (1 - \mathbf{x}^{a_1}) \cdots (1 - \mathbf{x}^{a_m})$ provided $hf = g$. For example, the series $f = \sum_{k=-\infty}^{+\infty} x^k$ in $\mathbb{Q}((x))$ is identified with zero, since $f(x)(1 - x) = 0$. In this approach, the valuation $\mathfrak{F}'$ can be viewed as an *algebra homomorphism*, provided the ring structure on $\mathbb{Q}((\mathbf{x}))$ is given by the Hadamard product $\star$:

$$f \star g = \sum_{m \in \mathbb{Z}^d} (a_m b_m) \mathbf{x}^m, \quad \text{where} \quad f = \sum_{m \in \mathbb{Z}^d} a_m \mathbf{x}^m \quad \text{and} \quad g = \sum_{m \in \mathbb{Z}^d} b_m \mathbf{x}^m.$$

4. Complexity of Generating Functions

Let $P \subset \mathbb{R}^d$ be a rational polyhedron. In this section, we show that if the dimension d is fixed, then the generating function $f(P;\mathbf{x})$ has a representation whose complexity is bounded by a polynomial in the input size of P. If P does not contain straight lines, then $f(P;\mathbf{x})$, "in principle", encodes all the information about the set $P \cap \mathbb{Z}^d$ of integer points in P. We claim that this information can be encoded in a compact way. In particular, our results imply that in any fixed dimension there is a polynomial time algorithm for counting integer points in a given rational polytope. First, we find a formula for $f(K;\mathbf{x})$, where K is a rational unimodular cone (with vertex not necessarily at the origin).

(4.1) Lemma. *Let*

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

where $c_1, \ldots, c_d$ is a basis of the integer lattice $\mathbb{Z}^d$ and $\beta_i \in \mathbb{Q}$ are rational numbers. Let $u_1, \ldots, u_d$ be the (negative) dual basis of $\mathbb{Z}^d$:

$$\langle u_i, c_j\rangle = \begin{cases} -1 & \textit{if } \ i = j \\ \ \ 0 & \textit{if } \ i \neq j. \end{cases}$$

Then

$$f(K;\mathbf{x}) = \mathbf{x}^v \prod_{i=1}^{d} \frac{1}{1-\mathbf{x}^{u_i}}, \quad \textit{where} \quad v = -\sum_{i=1}^{d} \lfloor \beta_i \rfloor u_i,$$

and $\lfloor\cdot\rfloor$ denotes the integer part of a number.

Proof. Let $K_0 = \operatorname{co}\{u_1, \ldots, u_d\}$. This is a unimodular cone with vertex at the origin, and by Example 3.3, we have

$$f(K_0;\mathbf{x}) = \prod_{i=1}^{d} \frac{1}{1-\mathbf{x}^{u_i}}.$$

A point $x \in \mathbb{R}^d$ is an integer point in K if and only if for $i = 1, \ldots, d$, the value $\langle c_i, x\rangle$ is an integer which does not exceed β_i, and hence does not exceed $\lfloor \beta_i \rfloor$. In other words, the set of integer points in K and in the translation $K_0 + v$ coincide. Therefore, $f(K, \mathbf{x}) = f(K_0 + v, \mathbf{x}) = \mathbf{x}^v f(K_0, x)$ (by Part 2 of Theorem 3.1), and the proof is complete. □

Remark. If we fix $c_1, \ldots, c_d$ and allow $\beta_1, \ldots, \beta_d$ to vary, the denominator of the rational function $f(K, \mathbf{x})$ does not change.

The following result is proved in [3].

(4.2) Theorem. *Let us fix d. There exists a polynomial time algorithm, which, given a rational polyhedral cone $K \subset \mathbb{R}^d$, computes unimodular cones $K_i : \ i \in I$ and numbers $\epsilon_i \in \{-1, 1\}$ such that*

$$[K] = \sum_{i \in I} \epsilon_i [K_i].$$

In particular, the number $|I|$ of cones in the decomposition is bounded by a polynomial in the input size of K.

Sketch of Proof. Using triangulation, we reduce the problem to the case of a simple cone $K = \mathrm{co}\{u_1, \dots, u_d\}$, where $u_1, \dots, u_d \subset \mathbb{Z}^d$ are linearly independent integer points (we leave aside lower-dimensional cones, which can be treated in a similar way). Let us introduce the *index* $\mathrm{ind}(K)$ which measures how far is K from being unimodular: $\mathrm{ind}(K) = |u_1 \wedge \dots \wedge u_d|$ is the volume of the parallelepiped spanned by the generators $u_1, \dots, u_d$. One can show that $\mathrm{ind}(K)$ is the number of integer points in the fundamental parallelepiped Π of K (cf. Example 3.3). Thus the index of a cone is a positive integer which equals 1 if and only if K is unimodular. One can show that $\log \mathrm{ind}(K)$ is bounded by a polynomial in the input size of K. We are going to iterate a procedure which replaces $[K]$ by a linear combination of $[K_j]$, where K_j are rational cones with smaller indices. Let us consider a parallelepiped:

$$B = \Big\{\alpha_1 u_1 + \dots + \alpha_d u_d : |\alpha_j| \le \big(\mathrm{ind}(K)\big)^{-\frac{1}{d}} \ : \ j = 1, \dots, d\Big\}.$$

We observe that B is centrally symmetric and that the volume of B is 2^d. Therefore, by the Minkowski convex body theorem (see, for example, [35]), there is a non-zero integer point $w \in B$ (such a vector can be constructed efficiently using, for example, integer programming in dimension d, cf. [58]). For $j = 1, \dots, d$, let

$$K_j = \mathrm{co}\{u_1, \dots, u_{j-1}, w, u_{j+1}, \dots, u_d\}.$$

If $w = \alpha_1 u_1 + \dots + \alpha_d u_d$, then

$$\mathrm{ind}(K_j) = |u_1 \wedge \dots \wedge u_{j-1} \wedge w \wedge u_{j+1} \wedge \dots \wedge u_d| = |\alpha_j||u_1 \wedge \dots \wedge u_{j-1} \wedge u_j \wedge u_{j+1} \wedge \dots \wedge u_d|$$

$$= |\alpha_j| \mathrm{ind}(K) \le \Big(\mathrm{ind}(K)\Big)^{\frac{d-1}{d}}.$$

Furthermore, there is a decomposition

$$[K] = \sum_{j \in J} \epsilon_j [K_j] + \sum_F \epsilon_F [F],$$

where F ranges over lower-dimensional faces of K_j, and $\epsilon_j, \epsilon_F \in \{-1, 1\}$. If we iterate this procedure, we observe that the indices of the cones involved decrease doubly exponentially, whereas the number of cones increases only exponentially. Therefore, iterating the procedure $O(\log \log \mathrm{ind}(K))$ times, we end up with a decomposition of K into a linear combination of unimodular cones, with the number of cones bounded by a polynomial in $\log(\mathrm{ind}(K))$. □

Remark. Triangulations are not enough. As is well-known, every rational cone can be triangulated into unimodular cones (see, for example, [22, Section 2.6]). However, to ensure polynomial time complexity, it is important to use signed decompositions; triangulations alone are not enough, as the following simple example shows.

Let $K = \mathrm{co}\{u_1, u_2\} \subset \mathbb{R}^2$ be the planar cone spanned by vectors $u_1 = (1, 0)$ and $u_2 = (1, n)$, where n is a positive integer. To triangulate K into unimodular cones, we have to draw a line through each point $(1, i)$, $1 \le i < n$. Thus the number of cones will grow linearly in n, that is, exponentially in the size $O(\log n)$ of the input. However, to write a signed decomposition we need only three unimodular cones: let $w = (0,1)$, $K_1 = \mathrm{co}\{w, u_1\}$, $K_2 = \mathrm{co}\{w, u_2\}$, and $K_3 = \mathrm{co}\{w\}$. Then K_1, K_2 and K_3 are unimodular cones, and $[K] = [K_1] - [K_2] + [K_3]$.

(4.3) *Remark. The duality trick.* Once we have a decomposition

$$[K] = \sum_{i \in I} \epsilon_i [K_i],$$

where K_i are unimodular cones, we can write

$$f(K; \mathbf{x}) = \sum_{i \in I} \epsilon_i f(K_i; \mathbf{x}) \tag{4.3.1}$$

(see Theorem 3.1). Since there are explicit formulas for $f(K_i; \mathbf{x})$ (see Example 3.3), we get an explicit formula for $f(K; \mathbf{x})$. The complexity of such a formula, as asserted by Theorem 4.2, is bounded by a polynomial in the input size of K. There is a trick which allows us to obtain a decomposition of the type (4.3.1), where all K_i are unimodular *and* d-dimensional. The idea can already be found in the seminal paper [6]. Sometimes this trick significantly reduces the computational complexity of the formula. Namely, let K^* be the dual cone to K. Let us apply the iterative procedure of Theorem 4.2 to K^*, discarding lower-dimensional cones on every step. Thus we get a decomposition

$$[K^*] = \sum_{i \in I} \epsilon_i [K_i] \qquad \text{modulo lower-dimensional cones,}$$

where the K_i are d-dimensional unimodular cones. The dual of a lower-dimensional cone contains a straight line. Therefore, by Corollary 2.8 we get that

$$[K] = \sum_{i \in I} \epsilon_i [K_i^*] \qquad \text{modulo cones with straight lines.}$$

From Theorem 3.1, we get that

$$f(K; \mathbf{x}) = \sum_{i \in I} \epsilon_i f(K_i^*; \mathbf{x}).$$

Note that the K_i^* are unimodular, provided that the K_i are unimodular. If the cone K is defined by linear inequalities in $\mathbb{R}^d$, then the complexity of the algorithm and the resulting formula is $\mathcal{L}^{O(d)}$, where $\mathcal{L}$ is the input size of K. The complexity of the algorithm of Theorem 4.2, as stated, can be as bad as $\mathcal{L}^{\Omega(d^2)}$ if K is given as a conic hull of integer vectors in $\mathbb{Z}^d$, and $\mathcal{L}^{\Omega(d^3)}$ if K is given by a set of linear inequalities. The savings in computational complexity comes from the fact that if we iterate the procedure of Theorem 4.2 as stated, the number of cones in every step grows by a factor of 2^d. If we discard lower-dimensional cones, the number of cones in every step grows by a factor of d only.

We are ready to state the main result of this section. We not only compute the expression for $f(P; \mathbf{x})$, but we also describe how it changes when the facets of P are moved parallel to themselves so that the combinatorial structure of P does not change.

(4.4) Theorem. *Let us fix d. There exists a polynomial time algorithm, which, for a given rational polyhedron $P \subset \mathbb{R}^d$,*

$$P = \Big\{x \in \mathbb{R}^d : \langle c_i, x\rangle \leq \beta_i : \ i = 1, \ldots, m\Big\}, \qquad \textit{where} \quad c_i \in \mathbb{Z}^d \quad \textit{and} \quad \beta_i \in \mathbb{Q}$$

computes the generating function $f(P; \mathbf{x}) = \sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$ in the form:

$$f(P; \mathbf{x}) = \sum_{i \in I} \epsilon_i \frac{\mathbf{x}^{a_i}}{(1 - \mathbf{x}^{b_{i1}}) \cdots (1 - \mathbf{x}^{b_{id}})}, \tag{4.4.1}$$

where $\epsilon_i \in \{-1, 1\}$, $a_i \in \mathbb{Z}^d$, and $b_{i1}, \ldots, b_{id}$ is a basis of $\mathbb{Z}^d$ for each i.

Suppose that the vectors $c_i : \ i = 1, \ldots, m$ are fixed and the β_i vary in such a way that the combinatorial structure of the polyhedron $P = P(\beta) : \ \beta = (\beta_1, \ldots, \beta_m)$ stays the same. Then the exponents b_{ij} in the denominator of each fraction remain the same, whereas the exponents $a_i = a_i(\beta)$ in the numerator change with $\beta \in \mathbb{Q}^m$ in the following way:

$$a_i = \sum_{j=1}^{d} \lfloor l_{ij}(\beta) \rfloor b_{ij}, \tag{4.4.2}$$

where $l_{ij} : \mathbb{Q}^m \longrightarrow \mathbb{Q}$ are linear functions and $\lfloor \cdot \rfloor$ is the integer part. If β is such that $P(\beta)$ is an integer polytope, then $l_{ij}(\beta) \in \mathbb{Z}$ for each pair i, j.

The computational complexity of the algorithm for finding (4.4.1) *and* (4.4.2) *is $\mathcal{L}^{O(d)}$, where $\mathcal{L}$ is the input size of P. In particular, the number $|I|$ of terms in* (4.4.1) *is $\mathcal{L}^{O(d)}$.*

Proof. Let $\operatorname{Vert}(P)$ be the set of vertices of P. For $v \in \operatorname{Vert}(P)$, let $I_v = \{i \in I : \langle c_i, v\rangle = \beta_i\}$ be the set of those inequalities that are active at v, and let $N(P, v) = \operatorname{co}\{c_i : \ i \in I_v\}$ be the conic hull of the normals of the facets containing v. Then for the tangent cone $\operatorname{cone}(P, v)$, we have

$$\operatorname{cone}(P, v) = -N^*(P, v) + v.$$

Using Theorem 4.2 and Remark 4.3, we construct d-dimensional unimodular cones $K(v, j) : \ j \in J_v$ and numbers $\epsilon(v, j) \in \{-1, 1\}$ such that

$$[N(P, v)] = \sum_{j \in J_v} \epsilon(v, j)[K(v, j)] \qquad \text{modulo lower-dimensional cones.}$$

Therefore, for the supporting cone of P at v, we have:

$$[\operatorname{cone}(P, v)] =$$

$$[v - N^*(P, v)] = \sum_{j \in J_v} \epsilon(v, j)[v - K^*(v, j)] \quad \text{modulo cones with straight lines.}$$

Thus,

$$f\big(\operatorname{cone}(P, v); \mathbf{x}\big) = \sum_{j \in J_v} \epsilon(v, j) f\big(v - K^*(v, j); \mathbf{x}\big)$$

by Part 4 of Theorem 3.1. Now, each cone $v - K^*(v, j)$ is a unimodular cone with vertex v. If the c_i are fixed, as long as the combinatorial structure of P does not change, each vertex $v = v(\beta)$ of P changes linearly with $\beta \in \mathbb{R}^m$. Therefore, we can apply Lemma 4.1 to find $f(v - K^*(v, j); \mathbf{x})$. Applying Theorem 3.5, we get

$$f(P; \mathbf{x}) = \sum_{v \in \mathrm{Vert}(P)} f\big(\mathrm{cone}(P, v); \mathbf{x}\big),$$

and the proof is complete. □

If $P \subset \mathbb{R}^d$ is a polytope, then

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

is a polynomial, whose expression as the sum of monomials can be very long. Theorem 4.4 asserts that if P is a rational polytope, then this polynomial can be written as a short rational function. A typical example is provided by a polynomial $\sum_{k=1}^n x^k$, containing n monomials, which can be written as the rational function $(1 - x^{n+1})/(1 - x)$, which one needs only $O(\log n)$ bits to write.

Theorem 4.4 implies that if the dimension d is fixed, then the valuation

$$\mathfrak{F} : \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{Q}(\mathbf{x})$$

is computable in polynomial time.

5. Efficient Counting of Lattice Points

If $P \subset \mathbb{R}^d$ is a rational polytope, then the generating function

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

is a polynomial, and its value at $\mathbf{x} = (1, \ldots, 1)$ is the number of integer points $|P \cap \mathbb{Z}^d|$ in P. Note that if we compute $f(P; \mathbf{x})$ as a short rational function, as provided by Theorem 4.4, then the point $\mathbf{x} = (1, \ldots, 1)$ is a pole of each fraction in the representation (4.4.1). This can be can be handled by taking an appropriate residue or by computing the value of $f(P; \mathbf{x})$ at a point $\mathbf{x}$ close to $\mathbf{1} = (1, \ldots, 1)$ and rounding the answer to the nearest integer as in [17]. We use the "residue" approach, suggested in [6] and also used in [3], where the first polynomial time algorithm for counting integer points in a rational polytope was constructed.

(5.1) Definition. Let us consider the function

$$F(\tau; \xi_1, \ldots, \xi_d) = \prod_{i=1}^d \frac{\tau \xi_i}{1 - \exp\{-\tau \xi_i\}}$$

in $d + 1$ complex variables τ and $\xi_1, \ldots, \xi_d$. It is easy to see that F is analytic in a neighborhood of the origin $\tau = \xi_1 = \ldots = \xi_d = 0$ and therefore there exists an expansion

$$F(\tau; \xi_1, \ldots, \xi_d) = \sum_{k=0}^{+\infty} \tau^k \mathrm{td}_k(\xi_1, \ldots, \xi_d),$$

where $\mathrm{td}_k(\xi_1, \ldots, \xi_d)$ is a homogeneous polynomial of degree k, called the k-th *Todd polynomial* in $\xi_1, \ldots, \xi_d$. It is easy to check that td_k is a symmetric polynomial with rational coefficients (cf.[25], or Section 5.3 of [22]).

(5.2) A Polynomial Time Algorithm for Counting Integer Points in Rational Polytopes when the Dimension is Fixed

Suppose that the dimension d is fixed and $P \subset \mathbb{R}^d$ is a rational polytope. Let us use Theorem 4.4 to compute

$$f(P;\mathbf{x}) = \sum_{i \in I} \epsilon_i \frac{\mathbf{x}^{a_i}}{(1-\mathbf{x}^{b_{i1}}) \cdots (1-\mathbf{x}^{b_{id}})}.$$

Let us construct a vector $l \in \mathbb{Z}^d$ such that $\langle l, b_{ij} \rangle \neq 0$ for each i and j. To construct such an l efficiently, let us consider the "moment curve" $g(\tau) = (1, \tau, \tau^2, \ldots, \tau^{d-1}) \in \mathbb{R}^d$. For each b_{ij}, the function $\langle g(\tau), b_{ij} \rangle$: $\tau \in \mathbb{R}$ is a non-zero polynomial of degree at most $d-1$ in τ, and thus this function has at most $d-1$ zeros. Therefore, we can select l from the set of integer vectors $\{g(0), g(1), \ldots, g(m)\}$, where $m = d(d-1)|I| + 1$.

Let $l = (\lambda_1, \ldots, \lambda_d)$. For $\tau > 0$, let $\mathbf{x}_\tau = \big(\exp\{\tau\lambda_1\}, \ldots, \exp\{\tau\lambda_d\}\big)$ and let $\xi_{ij} = \langle l, b_{ij} \rangle$ and $\eta_i = \langle l, a_i \rangle$. Then

$$|P \cap \mathbb{Z}^d| = \lim_{\tau \longrightarrow 0} f(P;\mathbf{x}_\tau) = \lim_{\tau \longrightarrow 0} \sum_{i \in I} \epsilon_i \frac{\exp\{\tau\langle l, a_i \rangle\}}{(1-\exp\{\tau\langle l, b_{i1}\rangle\}) \cdots (1-\exp\{\tau\langle l, b_{id}\rangle\})}$$

$$= \lim_{\tau \longrightarrow 0} \frac{1}{\tau^d} \sum_{i \in I} \epsilon_i \frac{\tau^d \exp\{\tau\eta_i\}}{(1-\exp\{\tau\xi_{i1}\}) \cdots (1-\exp\{\tau\xi_{id}\})}.$$

Now, each fraction

$$h_i(\tau) = \frac{\tau^d \exp\{\tau\eta_i\}}{(1-\exp\{\tau\xi_{i1}\}) \cdots (1-\exp\{\tau\xi_{id}\})}$$

is a holomorphic function in a neighborhood of $\tau = 0$ and the d-th coefficient of its Taylor series is

$$\frac{1}{\xi_{i1} \cdots \xi_{id}} \sum_{k=0}^{d} \frac{\eta_i^k}{k!} \operatorname{td}_{d-k}(\xi_{i1}, \ldots, \xi_{id}).$$

Finally, we get an efficient formula for the number of integer points in P:

$$(5.2.1) \qquad |P \cap \mathbb{Z}^d| = \sum_{i \in I} \frac{\epsilon_i}{\xi_{i1} \cdots \xi_{id}} \sum_{k=0}^{d} \frac{\eta_i^k}{k!} \operatorname{td}_{d-k}(\xi_{i1}, \ldots, \xi_{id}).$$

The construction of Algorithm 5.2 allows us to find out how the number of integer points in a rational polytope changes when the facets of the polytope are moved parallel to themselves, as long as the combinatorial type of the polytope does not change.

(5.3) Theorem. *Let us fix vectors $c_1, \ldots, c_m \in \mathbb{Z}^d$ such that for any $\beta \in \mathbb{Q}^m$: $\beta = (\beta_1, \ldots, \beta_m)$, the set*

$$P(\beta) = \Big\{ x \in \mathbb{R}^d : \langle c_i, x \rangle \leq \beta_i : \ i = 1, \ldots, m \Big\}$$

is a rational polytope in $\mathbb{R}^d$, if non-empty. Let $\mathcal{B} \subset \mathbb{Q}^m$ be a set such that for any $\beta_1, \beta_2 \in \mathcal{B}$, the polytopes $P(\beta_1)$ and $P(\beta_2)$ have the same combinatorial type. Then there exist linear functions $l_{ij} : \mathbb{Q}^m \longrightarrow \mathbb{Q}$ and rational numbers α_{ik} and γ_{ij} such that for any $\beta \in \mathcal{B}$,

$$|P(\beta) \cap \mathbb{Z}^d| = \sum_{i \in I} \sum_{k=1}^{d} \alpha_{ik} \left(\sum_{j=1}^{d} \gamma_{ij} \lfloor l_{ij}(\beta) \rfloor \right)^k, \tag{5.3.1}$$

where $\lfloor \cdot \rfloor$ is the integer part of a number. If for some $\beta \in \mathcal{B}$, $P(\beta)$ is an integer polytope, then $l_{ij}(\beta) \in \mathbb{Z}$. Furthermore, for any fixed dimension d, there is a polynomial time algorithm for computing the formula (5.3.1).

Proof. The theorem follows from Theorem 4.4 and Algorithm 5.2. Note that in (5.2.1), the numbers ξ_{ij} and ϵ_i do not change as long as $\beta \in \mathcal{B}$. Hence we let

$$\alpha_{ik} = \frac{\epsilon_i}{k! \xi_{i1} \cdots \xi_{id}} \operatorname{td}_{d-k}(\xi_{i1}, \ldots, \xi_{id}).$$

For $\beta \in \mathcal{B}$, we have $\eta_i = \eta_i(\beta) = \langle l, a_i(\beta) \rangle$ for some fixed $l \in \mathbb{Z}^d$, where by Theorem 4.4,

$$a_i(\beta) = \sum_{j=1}^{d} \lfloor l_{ij}(\beta) \rfloor b_{ij}$$

for some fixed $b_{ij} \in \mathbb{Z}^d$. Hence we let $\gamma_{ij} = \langle l, b_{ij} \rangle$. □

Remark. Relation to Integer Programming algorithms. Integer Programming is concerned with optimizing a given linear function on the set of integer points in a given rational polyhedron in $\mathbb{R}^d$. By using a standard trick of dichotomy, one can reduce an integer programming problem to a sequence of feasibility problems: given a rational polyhedron $P \subset \mathbb{R}^d$, decide whether P contains an integer point, and if so, find such a point. Integer Programming is difficult (NP-hard) in general, but it admits a polynomial time algorithm if the dimension d is fixed and not a part of the input. The first integer programming algorithm having polynomial time complexity in fixed dimension was constructed by H.W. Lenstra [38] (see [24], [40], [58] for survey and subsequent improvements). Of course, if we can count integer points in P, we can decide whether $P \cap \mathbb{Z}^d = \emptyset$. The catch is that in one of the crucial ingredients of Algorithm 5.2, namely in the proof of Theorem 4.2, we refer to integer programming in fixed dimension. Therefore, employing Algorithm 5.2 to solve an integer programming problem may seem to result in a vicious circle. However, as M. Dyer and R. Kannan show [17], one can avoid the dependence on Lenstra's algorithm in Theorem 4.2 by using an appropriate lattice reduction algorithm. Hence Algorithm 5.2 gives rise to a new linear programming algorithm, whose complexity is polynomial time when the dimension d is fixed and which uses lattice reduction (cf. [39]), but does not use "rounding" of a given convex body in $\mathbb{R}^d$, which is the second main ingredient in Lenstra's algorithm and its subsequent improvements (see [24], [40]). This rounding seems to be quite time-consuming and it would be interesting to find out if Algorithm 5.2 can compete with the known integer programming algorithms. On the other hand, Lenstra's algorithm can be naturally extended to "convex integer" problems, whereas Algorithm 5.2 heavily uses the polyhedral structure.

We conclude this section with a description of a very general "trick" which sometimes allows one to count lattice points efficiently even if the dimension is large.

Remark. Changing the lattice. Suppose that $\Lambda \subset \mathbb{Z}^d$ is a sublattice of a finite index $|\mathbb{Z}^d : \Lambda|$. Let Λ^* be the dual lattice. Therefore, $\mathbb{Z}^d \subset \Lambda^* \subset \mathbb{Q}^d$ and $|\Lambda^* : \mathbb{Z}^d| = |\mathbb{Z}^d : \Lambda|$.

For $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{C}^d$ and any vector $l = (\lambda_1, \ldots, \lambda_d) \in \mathbb{Q}^d$, let us define

$$e^{2\pi i l}\mathbf{x} = \big(\exp\{2\pi i \lambda_1\}x_1, \ldots, \exp\{2\pi i \lambda_d\}x_d\big).$$

Then the value of

$$\left(e^{2\pi i l}\mathbf{x}\right)^m = \exp\big\{2\pi i \langle l, m\rangle\big\}\mathbf{x}^m$$

depends only on the coset $\Lambda^* : \mathbb{Z}^d$ represented by l.

Furthermore, we have

$$\frac{1}{|\Lambda^* : \mathbb{Z}^d|} \sum_{l \in \Lambda^* : \mathbb{Z}^d} \left(e^{2\pi i l}\mathbf{x}\right)^m = \begin{cases} \mathbf{x}^m & \text{if } \ m \in \Lambda \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, if $P \subset \mathbb{R}^d$ is a rational polytope with a known function $f(P; \mathbf{x})$, one can compute the modified function

$$f(P, \Lambda; \mathbf{x}) = \sum_{m \in P \cap \Lambda} \mathbf{x}^m$$

by the formula

$$f(P, \Lambda; \mathbf{x}) = \frac{1}{|\Lambda^* : \mathbb{Z}^d|} \sum_{l \in \Lambda^* : \mathbb{Z}^d} f(P; e^{2\pi i l}\mathbf{x}).$$

Therefore, if the function $f(P; \mathbf{x})$ is known and the index $|\mathbb{Z}^d : \Lambda|$ is not very large, one can compute $f(P, \Lambda; \mathbf{x})$ efficiently.

The computational complexity of the algorithm above is exponential in the input size of Λ even when the dimension d is fixed, but for lattices Λ of small index $|\mathbb{Z}^d : \Lambda|$, it may appear computationally useful. Suppose, for example, that $P = [0, n_1] \times [0, n_2] \times \ldots \times [0, n_d]$ is the d-dimensional integer "box". Then there is an explicit formula for the generating function $f(P; \mathbf{x})$:

$$f(P; \mathbf{x}) = \prod_{i=1}^{d} \frac{1 - x_i^{n_i+1}}{1 - x_i}.$$

Thus there is an algorithm for counting points in $P \cap \Lambda$ whose complexity is polynomial in the index $|\mathbb{Z}^d : \Lambda|$ (d need not be fixed).

This approach was used in [9] and in [2] in a particular situation.

This construction can be dualized. Suppose that $\Lambda \subset \mathbb{Q}^d$ is a lattice such that $\mathbb{Z}^d \subset \Lambda$. Now $m \in \Lambda$ no longer needs to be integer vector, so in order to resolve the ambiguity of $\mathbf{x}^m$, it is convenient to make the substitution $x_j = e^{\lambda_j}$, $j = 1, \ldots, d$ and hence interpret the monomial $\mathbf{x}^m$ as the function $\mathbb{C}^d \longrightarrow \mathbb{C}$, $l \longmapsto \exp\big\{\langle l, m\rangle\big\}$, where $l = (\lambda_1, \ldots, \lambda_d)$.

Then

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

(Clearly, the sum does not depend on a particular choice of coset representatives l). Therefore, if $f(P; \mathbf{x})$ is known and the index $|\Lambda : \mathbb{Z}^d|$ is small, then one can compute $f(P, \Lambda; \mathbf{x})$ efficiently. We can iterate the two constructions: first we take a sublattice $\Lambda_1 \subset \mathbb{Z}^d$ of a small index, then a superlattice $\Lambda_2 \supset \Lambda_1$ of a small index, then a sublattice $\Lambda_3 \subset \Lambda_2$, and so forth.

6. Existence of "Local Formulae"

The results of Sections 4 and 5 provide a satisfactory solution of the counting problem when the dimension of the ambient space is fixed. If the dimension d is allowed to grow, the algorithms can become less efficient than straightforward enumeration. If the dimension is allowed to grow, the problem of "efficient counting" seems to be ill-posed, since much depends on the particulars of the polytope. For example, it becomes relevant whether the polytope is given by the list of its vertices or by the list of its facets. In this section, we explain our approach to what the "right" counting problem is when the dimension is allowed to grow.

P. McMullen (see [45]) proved that the number of integer points in a rational polytope P can be expressed as a linear combination of the volumes of the faces of the polytope, where the coefficient of $\operatorname{vol}(F)$, where F is a face of P, depends only on the translation class mod $\mathbb{Z}^d$ of the supporting cone $\operatorname{cone}(P, F)$ of P at F (see Definition 3.4).

(6.1) Theorem ("Local Formula"). *For every rational cone $K \subset \mathbb{R}^d$ one can define a rational number $\phi(K)$, such that*

(1) *The function ϕ is invariant under lattice translations:*

$$\phi(K) = \phi(K + u) \quad \textit{for any} \quad u \in \mathbb{Z}^d;$$

(2) *For any rational polytope $P \subset \mathbb{R}^d$,*

$$|P \cap \mathbb{Z}^d| = \sum_F \operatorname{vol}(F) \phi\big(\operatorname{cone}(P, F)\big),$$

where the sum is taken over all faces F of P, and $\operatorname{vol}(F)$ *is the volume of F, measured intrinsically in its affine span.*

Theorem 6.1 immediately implies that the number of integer points $|kP \cap \mathbb{Z}^d|$ in the dilated polytope kP, where k is positive integer, is a quasipolynomial $\sum_{i=1}^d f_i(k) k^d$, where $f_i(k)$ are periodic functions. If P is integral, then we get a genuine polynomial, namely the Ehrhart polynomial of P, see Example 1.5.

The function ϕ fails badly to be unique. Essentially, one can get the existence of ϕ via Hahn-Banach type reasoning (cf. [44]). In the next section, we sketch a more constructive approach, also due to McMullen, via what we call the "Combinatorial Stokes Formula" (the formula belongs to McMullen, whereas its name is our invention). In many important cases, ϕ can be chosen to be a valuation on rational cones.

Note, if dim $P = d$ and dim $F = k$, then the apex of cone(P, F) is a k-dimensional affine subspace. Thus the cone is just the product of a $(d-k)$-dimensional pointed cone with the apex at the origin and a rational subspace. Hence cone(P, F) looks "simple" when $d - k$ is small. In our opinion, the "right" problem to consider is the following:

6.2 Problem

If k is fixed and d is allowed to grow, find a computationally efficient choice of ϕ on rational cones with $(d-k)$-dimensional apex.

In other words, we are interested in computing the highest terms of the expression for the Ehrhart (quasi)polynomial of P. Problem 6.2 is still not solved completely. In Section 8, we will see this problem solved for integer polytopes: computation of any fixed number of the highest coefficients of the Ehrhart polynomial of an integer polytope reduces in polynomial time to computation of volumes of faces (see [4]). However, the problem of finding computable functions ϕ for rational polytope is not yet solved.

The supporting cone cone(P, F) is "unnecessarily large", as it contains an affine subspace. Sometimes it is desirable to get a more "condensed" representation for the function ϕ.

(6.3) Definition. Let $P \subset \mathbb{R}^d$ be a full-dimensional polyhedron and let $F \subset P$ be a face. The *normal cone* $N(P, F)$ is the cone spanned by the outer normals of facets of P that contain F. In other words, if

$$P = \{x \in P : \langle c_i, x\rangle \leq \beta_i : i = 1, \ldots, m\},$$

F is a face of P, and $I_F = \{i : \langle c_i, x\rangle = \beta_i \text{ for every } x \in F\}$ is the set of inequalities that are active on F, then

$$N(P, F) = \operatorname{co}\{c_i : i \in I_F\}.$$

If dim $F = k$, then $N(P, F)$ is a $(d-k)$-dimensional pointed cone with vertex at the origin.

Thus, for a full-dimensional integer polytope $P \subset \mathbb{R}^d$, Theorem 6.1 asserts that

$$|P \cap \mathbb{Z}^d| = \sum_F \phi\big(N(P, F)\big)\operatorname{vol}(F),$$

where ϕ is a function on pointed rational cones. The values of ϕ on lower-dimensional cones determine higher-dimensional coefficients of the Ehrhart polynomial. If P is a rational polytope, which is not an integer polytope, the value of ϕ depends not only on the normal cone $N(P, F)$, but also on the translation class mod $\mathbb{Z}^d$ of the affine hull aff(F) of the face F.

7. Combinatorial Stokes' Formula and Its Applications

Roughly speaking, the main idea of this section is to express the number $|P \cap \mathbb{Z}^d|$ of integer points in a polytope P as a sum of the "main term", which is the volume of P, and the "correction term," which is associated with the boundary ∂P of P. Naturally, we will have to consider lower-dimensional sublattices of $\mathbb{Z}^d$. This explains why it is convenient to consider right from the beginning a general lattice $\Lambda \subset \mathbb{R}^d$ (discrete additive subgroup of $\mathbb{R}^d$), rather than only $\mathbb{Z}^d$.

(7.1) Definitions. Let us fix a lattice $\Lambda \subset \mathbb{R}^d$. A polytope $P \subset \mathbb{R}^d$ is called a *lattice* or Λ-polytope provided that its vertices belong to Λ. A polytope $P \subset \mathbb{R}^d$ is called a *rational* or $\mathbb{Q}\Lambda$-polytope provided that for some positive integer n, $nP = \{nx : x \in P\}$ is a lattice polytope. Let $\mathcal{P}_c(\mathbb{Q}\Lambda)$ be the subspace (subalgebra) of $\mathcal{P}_c(\mathbb{R}^d)$ spanned by the indicator functions $[P]$ of rational polytopes. A valuation $\Phi : \mathcal{P}_c(\mathbb{Q}\Lambda) \longrightarrow V$ is called *simple* provided $\Phi([P]) = 0$ when $\dim\ P < d$. A valuation Φ is called Λ-*invariant* provided $\Phi([P + \lambda]) = \Phi([P])$ for any $\lambda \in \Lambda$. A valuation Φ is called *centrally symmetric* provided $\Phi([P]) = \Phi([-P])$ for every polytope P. Often we write $\Phi(P)$ instead of $\Phi([P])$.

(7.2) Cones and angles

Let $K \subset \mathbb{R}^d$ be a cone. Let S^{d-1} be the unit sphere centered at the apex of K. We define $\alpha_d(K)$ to be the *spherical measure* of the intersection $K \cap S^{d-1}$ normalized in such a way that the spherical measure of the whole sphere S^{d-1} is 1. We also agree that if $d = 0$, then $\alpha_0(0) = 1$. Clearly, $\alpha_d(K) = 0$ if K is not a full-dimensional cone. The *intrinsic measure* $\alpha(K)$ is defined as the spherical measure $\alpha_k(K)$ in the affine hull of K, where $k = \dim\ K$.

Finally, let $P \subset \mathbb{R}^d$ be a d-dimensional polyhedron, and let $F \subset K$ be a k-dimensional face of P. The *exterior angle* $\gamma(P, F)$ of P at F is the intrinsic measure of the normal cone of P at F (see Definition 6.3); that is, $\gamma(P, F) = \alpha\big(N(P, F)\big)$.

If $K \subset \mathbb{R}^d$ is a d-dimensional polyhedral cone, then

$$\sum_F \alpha(F)\gamma(K, F) = 1, \tag{7.2.1}$$

where the sum is taken over all non-empty faces F of K (see [41]). The proof, also due to McMullen, is immediate: for every point $x \in \mathbb{R}^d$, let $n(x) \in K$ be the (unique) point closest to x in the Euclidean metric. Then the summands of (7.2.1) correspond to a dissection of $\mathbb{R}^d$ into pieces, each of which consists of those points x such that $n(x)$ is in the relative interior of a given face F.

An important example of a simple lattice-invariant valuation related to lattice point counting arises when we count lattice points with their "solid angles."

(7.3) The solid angle valuation ρ. For a polyhedron $P \subset \mathbb{R}^d$ and a point $x \in \mathbb{R}^d$, let us define the *solid angle $\beta(P, x)$ of P at x* in the following way: let $B_r(x)$ denote the ball centered at x of radius r. We let

$$\beta(P, x) = \lim_{r \longrightarrow 0} \frac{\mathrm{vol}(P \cap B_r(x))}{\mathrm{vol}(B_r(x))},$$

where "vol" is the usual volume in $\mathbb{R}^d$. For example, if $x \notin P$ then $\beta(P, x) = 0$. Similarly, if $\dim\ P < d$ then $\beta(P, x) = 0$ for any x. Furthermore, $\beta(P, x) = 1$ if and only if x is in the interior of P. If P is d-dimensional and x lies in the relative interior of a facet of P, then $\beta(P, x) = 1/2$. Generally, if x lies in the relative interior of a face F of P, then $\beta(P, x)$ is the spherical measure of the supporting cone of P at F.

Let

$$\rho(P) = \sum_{x \in \Lambda} \beta(P, x).$$

It is easy to see that ρ extends to a simple centrally-symmetric Λ-invariant valuation

$$\rho : \mathcal{P}_c(\mathbb{Q}\Lambda) \longrightarrow \mathbb{R}.$$

From (7.2.1), one can deduce that

$$|P \cap \Lambda| = \sum_F \rho(F)\gamma(P, F), \tag{7.3.1}$$

where the sum is taken over all faces F of P and $\rho(F)$ is defined intrinsically in the affine span of F.

Let us fix a lattice $\Lambda \subset \mathbb{R}^d$, with rank $\Lambda = d$. A subspace $L \subset \mathbb{R}^d$ is called a *lattice subspace*, if it is spanned by lattice points. Similarly, an affine subspace $A \subset \mathbb{R}^d$ is called a *lattice subspace* provided it is a lattice translation of a linear lattice subspace. A *rational* subspace is a translation of a lattice linear subspace by a vector in $\mathbb{Q}\Lambda$.

Suppose we are given a simple Λ-invariant valuation Φ. The idea of the "Combinatorial Stokes Formula" is to construct a family of valuations $\{\phi_A\}$ concentrated on affine rational hyperplanes $A \subset \mathbb{R}^d$. It turns out that each valuation ϕ_A is simple, considered as a valuation in the $(d-1)$-dimensional space A. If $F \subset A$ is a polytope, instead of writing $\phi_A(F)$, we write simply $\phi(F)$. Our theorem follows [42].

(7.4) Theorem. *Let us fix a lattice $\Lambda \subset \mathbb{R}^d$ and a simple lattice invariant valuation Φ on rational polytopes in $\mathbb{R}^d$. Then there exist*

a number $\alpha \in \mathbb{R}$,
a function $\kappa : \mathbb{R}^d \longrightarrow \mathbb{R}$, and
a family of simple valuations $\{\phi_A\}$, associated with rational hyperplanes $A \subset \mathbb{R}^d$,

such that the following properties are satisfied:

1. *valuations ϕ_A are Λ-invariant: if P and Q are rational $(d-1)$-dimensional polytopes and P is a lattice translation of Q, then $\phi(P) = \phi(Q)$;*
2. *if $P \subset \mathbb{R}^d$ is a rational polytope, then*

$$\Phi(P) = \alpha \operatorname{vol}(P) + \sum_F \kappa(n_F)\phi(F),$$

where the sum is taken over all facets of P and n_F is the outer unit normal to F;

3. *κ is an odd function: $\kappa(-u) = -\kappa(u)$ for each $\in \mathbb{R}^d$.*

Sketch of Proof. Without loss of generality, we assume that $\Lambda = \mathbb{Z}^d$. We proceed by induction on d. For $d = 0$, the result is clear.

Suppose that $d \geq 1$. Let us consider $\mathbb{R}^{d-1}$ as a hyperplane in $\mathbb{R}^d$ (the last coordinate is 0). Let us define a valuation Ψ on $\mathbb{R}^{d-1}$ by $\Psi(Q) = \Phi(Q \times [0,1])$. Clearly, Ψ is a simple $\mathbb{Z}^{d-1}$-invariant valuation in $\mathbb{R}^{d-1}$, so we can apply the induction hypothesis to Ψ. Let $\alpha = \alpha_\Psi$ be the corresponding number, $\kappa = \kappa_\Psi : \mathbb{R}^{d-1} \longrightarrow \mathbb{R}$ be the corresponding function, and let $\{\psi_H\}$ be the corresponding family of valuations for rational hyperplanes $H \subset \mathbb{R}^{d-1}$. If $H \subset \mathbb{R}^{d-1}$ is a rational hyperplane, then $A = H \oplus \mathbb{R}^1$ is a rational hyperplane in $\mathbb{R}^d$ whose normal vector is in $\mathbb{R}^{d-1}$. Using the induction hypothesis, one can show that the valuations ψ_H can be extended to valuations ϕ_A in such a way that $\phi_A(Q \times [0,1]) = \psi_H(Q)$ for any rational polytope $Q \in H$. Hence we have constructed valuations ϕ_A on the hyperplanes whose normal vectors are in $\mathbb{R}^{d-1}$. For a rational polytope $P \subset \mathbb{R}^d$, let

$$\overline{\Phi}(P) = \Phi(P) - \alpha \mathrm{vol}_d(P) - \sum_F \kappa(n_F)\phi(F),$$

where the sum is taken over all facets of P whose normal vectors are in $\mathbb{R}^{d-1}$. Using Theorem 2.9, one can show that $\overline{\Phi}$ is in fact a simple Λ-invariant valuation.

It is clear that $\overline{\Phi}(Q \times [0,1]) = 0$, where Q is a lattice polytope in $\mathbb{R}^{d-1}$. Since $\overline{\Phi}$ is a simple valuation, we conclude that $\overline{\Phi}(P) = 0$ if P is a "lattice prism" $Q \times [m,n]$, where $m, n \in \mathbb{Z}$.

Now we are ready to define ϕ_A for any rational hyperplane $A \subset \mathbb{R}^d$ and κ for all $u \in \mathbb{R}^d$. Let $A \subset \mathbb{R}^d$ be a rational affine hyperplane and let $Q \subset A$ be a rational polytope. Translating by a lattice vector, if necessary, we can always assume that Q is in the "upper halfspace" of $\mathbb{R}^d$ (the last coordinate is positive). Let Q' be the projection of Q "down" onto $\mathbb{R}^{d-1}$, and let $\Pi(Q) = \mathrm{conv}(Q \cup Q')$ be the "skewed prism" with "bottom" facet Q' and "top" facet Q. Let

$$\phi_A(Q) = \overline{\Phi}(\Pi(Q)).$$

One can show that $\phi_A(Q)$ is well-defined and that the family $\{\phi_A\}$ is Λ-invariant, since if we choose a lattice translation P of Q, the difference between the values of $\overline{\Phi}$ on the "skewed prisms" $\Pi(Q)$ and $\Pi(P)$ will be the value of $\overline{\Phi}$ on a "right prism" $Q \times [m,n]$, which is zero. Let u_d denote the last coordinate of a vector $u \in \mathbb{R}^d$. If $u_d = 0$ we can think of u as a vector in $\mathbb{R}^{d-1}$. Let us define κ by

$$\kappa(u) = \begin{cases} 1 & \text{if} \quad u_d > 0 \\ -1 & \text{if} \quad u_d < 0 \\ \kappa_\Psi(u) & \text{if} \quad u_d = 0. \end{cases}$$

Now the theorem follows since

$$[P] = \sum_F \kappa(n_F)[\Pi(F)] \qquad \text{modulo lower-dimensional polytopes,}$$

where the sum is taken over all facets F of P. □

The following example justifies, in our opinion, the name "Combinatorial Stokes' Formula" for Theorem 7.4.

(7.5) Example. Exponential Valuations. Let Λ^* be the lattice dual to Λ:

$$\Lambda^* = \{x \in \mathbb{R}^d : \langle x, u \rangle \in \mathbb{Z} \quad \text{for each} \quad u \in \Lambda\}.$$

Then to each $l \in \Lambda^*$ one can associate a simple Λ-invariant valuation Φ_l defined by

$$\Phi_l(P) = \int_P \exp\{2\pi i \langle l, x \rangle\}\, dx.$$

It is clear that if $l = 0$ then $\Phi_l(P) = \mathrm{vol}(P)$. If $l \neq 0$ then (the ordinary) Stokes formula implies that

$$\int_P \exp\{2\pi i \langle l, x\rangle\}\ dx = \frac{1}{\langle l, c\rangle} \sum_F \langle c, n_F\rangle \int_F \exp\{2\pi i \langle l, x\rangle\}\ dx_F,$$

where c is any vector such that $\langle l, c\rangle \neq 0$. Here, the sum is taken over all facets F of P, and dx_F is Lebesgue measure on the supporting hyperplane of F (see [2]). Therefore, Theorem 7.4 holds with $\kappa(u) = \langle c, u\rangle / \langle l, c\rangle$ and $\{\phi_A\}$ being a family of exponential valuations on rational hyperplanes in $\mathbb{R}^d$.

Theorem 7.4 has an interesting corollary. We say that a set $X \subset \mathbb{R}^d$ is centrally symmetric provided there is a point y such that $2y - x \in X$ for any $x \in X$.

(7.6) Corollary. *Let Φ be a Λ-invariant simple centrally symmetric valuation. There exists a constant α such that for each lattice polytope P whose facets are centrally symmetric, one has*

$$\Phi(P) = \alpha \cdot \mathrm{vol}\ \ P.$$

Proof. We have

$$\Phi(P) = \frac{\Phi(P) + \Phi(-P)}{2}.$$

Expressing $\Phi(P)$ and $\Phi(-P)$ by Theorem 7.4, we notice that all the terms except the main one cancel each other out. □

Applications to the solid angle valuation ρ

Let us consider the valuation ρ of Example 7.3, which counts every lattice point in P with weight equal to the solid angle at that point.

(7.7) Corollary. *Let $\Lambda \subset \mathbb{R}^d$ be a lattice of rank d and let $P \subset \mathbb{R}^d$ be a lattice polytope whose facets are centrally symmetric. Then*

$$\rho(P) = \frac{\mathrm{vol}(P)}{\det \Lambda}.$$

Proof. By Corollary 7.6, it follows that $\rho(P) = \alpha \mathrm{vol}(P)$ for some α. Let $nP = \{nx : x \in P\}$ be a dilatation of P. Then

$$\alpha = \lim_{n \longrightarrow +\infty} \frac{\rho(nP)}{\mathrm{vol}(nP)} = \frac{1}{\det \Lambda}.$$

□

Corollary 7.7 can be considered as a “101-st” generalization of Pick’s formula (Example 1.2). Indeed, all facets of a polygon are centrally symmetric. Pick’s formula is equivalent to saying that if we count every integer point in a polygon P with weight equal to the angle at this point, we get the area of the polygon.

An interesting example of a polytope with centrally symmetric faces is provided by a *zonotope*, that is, the Minkowski sum of finitely many lattice intervals. To remove various normalizing factors, it is convenient to measure volumes of polytopes intrinsically, with respect to a given lattice. Namely, let us fix a lattice $\Lambda \subset \mathbb{R}^d$. Suppose that $P \subset \mathbb{R}^d$ is a lattice polytope and suppose that $k = \dim\ \ P$. Without loss of generality, we may assume that the affine hull A of P contains the origin. Then $\Lambda_A = \Lambda \cap A$ is a lattice of rank k, and we normalize the volume form in A in such a way that $\det \Lambda_A = 1$

(7.8) Corollary. *Let $P \subset \mathbb{R}^d$ be an integer zonotope, that is, the Minkowski sum of finitely many integer intervals. Then the number of integer points in P is expressed by the formula:*

$$|P \cap \mathbb{Z}^d| = \sum_F \operatorname{vol}(F)\gamma(P, F),$$

where the sum is taken over all faces F of P, $\gamma(P, F)$ is the exterior angle of P at F, and the volume of a face is measured intrinsically with respect to the lattice.

Proof. Since every face of a zonotope is centrally symmetric, the result follows from Corollary 7.7 and formula (7.3.1). Of course, it is quite easy to find a simple alternative proof which does not use Theorem 7.4. For a full-dimensional lattice parallelepiped Π (that is, the Minkowski sum of d linearly independent intervals) we have $\rho(\Pi) = \operatorname{vol}(\Pi)$, since lattice translates of Π tile the space $\mathbb{R}^d$ and both the volume and the valuation ρ are simple and lattice-invariant. Since a lattice zonotope can be dissected into lattice parallelepipeds (see, for example, [64, Lecture 7]), we get $\rho(P) = \operatorname{vol}(P)$ for any lattice zonotope P. Since every face of a lattice zonotope is a lattice zonotope itself, we use (7.3.1) to complete the proof. □

The boundary of a convex polytope can be represented as a union of lower-dimensional polytopes. Therefore, we can apply the Stokes formula of Theorem 7.4 recursively, first to the polytope P, then to its facets, then to its ridges, and so forth. We will end up with a decomposition involving volumes of faces and some "local" functions, depending only on the supporting cones at the faces (see [42]). Thus using Theorem 7.4, one can prove Theorem 6.1 first for the simple valuation ρ, and then, applying (7.3.1), for the number of integer points.

(7.9) Coefficients of the Ehrhart polynomial. For any rational polytope $P \subset \mathbb{R}^d$, there is a positive integer m such that the dilatation $Q_m = mP$ is an integer polytope. Let us consider the Ehrhart polynomial of Q_m (see Example 1.5)

$$|nQ_m \cap \mathbb{Z}^d| = a_d(Q_m)n^d + \ldots + a_0(Q_m),$$

and let us define $a_k(P) = a_k(Q_m)/m^k$ for $k = 0, \ldots, d$. It is easy to see that the numbers $a_k(P)$ are well-defined, that is independent on the choice of the scaling factor m. Furthermore, one can see that $a_k : \mathcal{P}_c(\mathbb{Q}^d) \longrightarrow \mathbb{Q}$ is a valuation. As U. Betke and M. Kneser proved [5], the coefficients a_k constitute a basis of the vector space of all valuations $\mathcal{P}_c(\mathbb{Q}^d) \longrightarrow \mathbb{Q}$ that are invariant under the affine transformations of $\mathbb{R}^d$ that map the lattice $\mathbb{Z}^d$ onto itself. Obviously, $a_k(P)$ are homogeneous, that is $a_k(rP) = r^k a_k(P)$, where $r > 0$ is a rational number. Corollary 7.8 implies that if P is a rational zonotope, then

$$a_k(P) = \sum_{F : \dim F = k} \operatorname{vol}(F)\gamma(P, F).$$

In particular, the coefficients of the Ehrhart polynomial of a zonotope are always non-negative. Curiously, if $k = 0, d-1$, or d, the formula holds true for *any* rational polytope P. One can show that in fixed dimension, one can compute the spherical measure of a given polyhedral cone within any given error $\epsilon > 0$ in polynomial time. Furthermore, even if the dimension is allowed to be a part of the input, using the technique of [16] (see [29] for recent improvements), one can come up with a randomized algorithm which, given an $\epsilon > 0$, approximates the spherical measure of a given polyhedral cone with relative error ϵ in time which is polynomial in the input size and ϵ^{-1}. The cone may be given as the convex hull of rays or as the intersection of subspaces. Thus in the class of integer zonotopes, Problem 6.2 has a satisfactory solution.

For general rational polytopes $P \subset \mathbb{R}^d$, Theorem 6.1 asserts that

$$a_k(P) = \sum_{F: \dim F = k} \operatorname{vol}(F) \phi\big(N(P, F)\big),$$

where ϕ is "some function" on the normal cones $N(P, F)$ of P at the k-dimensional faces F. There are 3-dimensional integer polytopes P, such that $a_1(P) < 0$. (Example: the tetrahedron $\Delta \subset \mathbb{R}^3$ with the vertices $(0, 0, 0), (1, 0, 0), (0, 1, 0)$ and $(1, 1, 13)$, see Section 5.3 of [22].) Hence, in general, ϕ can hardly be any geometric measure, it must reflect arithmetic structure. Computationally efficient choices of ϕ are discussed further in Sections 8 and 9.

Let $P \subset \mathbb{R}^d$ be an integer polytope. Instead of considering the coefficients $a_k(P)$, it can be useful to consider their linear combinations, $h_0^*(P), \ldots, h_d^*(P)$, defined by the following expression of a power series in one variable x as a rational function:

$$\sum_{n=0}^{\infty} \Big(a_0(P) + a_1(P)n + \ldots + a_d(P)n^d\Big) x^n = \frac{h_0^*(P) + h_1^*(P)x + \ldots + h_d^*(P)x^d}{(1-x)^{d+1}}.$$

Unlike the coefficients $a_k(P)$, the numbers $h_k^*(P)$ are monotone, that is $h_k^*(P) \geq h_k^*(Q)$ provided $Q \subset P$ are integer polytopes [60]. In particular, $h_k^*(P)$ are always non-negative. However, $h_k^*(P)$ are not homogeneous.

From the proof of Theorem 7.4, it is not at all clear what the functions ϕ in Theorem 6.1 might look like. The problem appears somewhat easier for exponential valuations Φ_l (Example 7.5). Since the main term is 0 unless $l = 0$ we conclude that if the highest term of the Ehrhart (quasi)polynomial of the valuation Φ_l is k, then l is orthogonal to some k-dimensional face of P.

A possible approach would be to relate the exponential valuations Φ_l of Example 7.5 and the solid angle valuation ρ (see Section 7.3). Let

$$\theta_\tau(x) = \tau^{d/2} \sum_{u \in \Lambda} \exp\big\{-\tau\pi \|x - u\|^2\big\}$$

$$= (\det \Lambda)^{-1} \sum_{l \in \Lambda^*} \exp\big\{-\pi \|l\|^2/\tau\big\} \exp\big\{2\pi i \langle l, x\rangle\big\},$$

where $\tau > 0$ is a parameter. It is then easy to show (see [1]) that

$$\lim_{\tau \longrightarrow +\infty} \int_P \theta_\tau(x)\, dx = \rho(P).$$

Thus we get a decomposition of ρ into the Fourier series of the valuations Φ_l:

$$\rho(P) = \lim_{\tau \longrightarrow +\infty} \sum_{l \in \Lambda^*} \exp\big\{-\pi \|l\|^2/\tau\big\} \Phi_l(P).$$

This approach was discovered independently and is being developed by R. Diaz and S. Robins [14].

8. Using Algebraic Geometry to Count Lattice Points

In recent years, many authors ([7], [10], [23], [31], [26], [49] [50], [53], [54], [55], [33]) have studied the problem of lattice point enumeration using the subject of toric varieties. In this section, we describe some of these results. In particular, we show how the valuation $\mathfrak{F}$ introduced in Section 3 plays a key role in various formulas for counting lattice points and for the closely-related problem of finding the Todd class of a toric variety.

Toric varieties, though very special and somewhat simple from the point of view of algebraic geometry, provide a powerful link between the theory of lattice polytopes and algebraic geometry. Early researchers in the field of toric varieties realized that finding a formula for a toric variety's *Todd class*, a characteristic class living in homology, would yield a formula for the number of lattice points in an integral polytope. Details of this connection, a direct result of the Riemann-Roch Theorem, are sketched following the statement of Algorithm 8.7. For a more complete discussion, see [22, Section 5.3]. This connection between polytopes and toric varieties has been very fruitful, especially over the last ten years. Indeed, much of the recent progress in counting lattice points is the result of a better understanding of the Todd class of a toric variety.

Interestingly, many of the lattice point formulas arising from the theory of toric varieties were later found to have elementary proofs, completely independent of algebraic geometry. A good example of this is Brion's Theorem (Theorem 3.5 above), which Brion first proved by applying a localization theorem for equivariant K-theory to toric varieties. However, we have seen in Section 3 (as have others before us) that it is not difficult to give an elementary proof of this simple and beautiful formula. Another example is Brion-Vergne's generalization of Khovanskii-Pukhlikov's Euler-Maclaurin summation formula for polytopes. Again, toric varieties provided motivation and an initial line of attack for these authors, though the final proofs in [8] are entirely elementary. One is left wondering what stopped these beautiful formulas from appearing long ago.

Let K be a d-dimensional rational, simple cone in $\mathbb{R}^d$. Here *simple* means that K is the convex hull of d rational rays from the origin. We denote the rays of K by $v_1, \ldots, v_d$, and identify each ray v_i with the first nonzero lattice point on that ray. By abuse of notation, we will use K to denote the d-by-d matrix whose ith row is v_i. Let K^* denote the dual cone, introduced in Section 2. Finally, if $\mathbf{x} = (x_1, \ldots, x_d)$, the following abbreviation will be convenient:

$$e^{\mathbf{x}} = (e^{x_1}, \ldots, e^{x_n}).$$

We are now ready to introduce a certain Laurent series $\mathfrak{s}_K$ in d variables $\mathbf{y} = (y_1, \ldots, y_d)$ corresponding to the rays of K. This Laurent series is a reparametrization of the function f of Theorem 3.1:

(8.1) Definition. To any d-dimensional rational, simple cone K in $\mathbb{R}^d$, we define

$$\mathfrak{s}_K(\mathbf{y}) = f(K^*; e^{-\mathbf{y}K}),$$

which represents a rational function in the variables $e^{y_1}, \ldots, e^{y_n}$. It will also be convenient to have a modified version $\mathfrak{t}_K$ of $\mathfrak{s}_K$. We define

$$\mathfrak{t}_K(\mathbf{y}) = \operatorname{ind}(K) y_1 y_2 \cdots y_d \cdot \mathfrak{s}_K(\mathbf{y}).$$

It is precisely this variant $\mathfrak{t}_K$ of the function f that will express for us the Todd class of a simplicial toric variety: we shall shortly see that $\mathfrak{t}_K$ defines a *power series* in the variables $y_1, \ldots, y_n$, and we can express the Todd class by taking the variables y_i in $\mathfrak{t}_K$, which correspond to rays of K, and substituting the corresponding divisor class. Before stating this precisely, we make some easy observations about $\mathfrak{s}_K$.

First notice that the expression for $\mathfrak{s}_K$ alters the function f of Theorem 3.1 in two ways: exponentials are substituted for the variables, and a linear change of coordinates is made. The linear change of coordinates has the pleasant effect of making $\mathfrak{s}_K$ invariant under lattice automorphisms. This is expressed in the following proposition, along with two other important properties of $\mathfrak{s}_K$.

(8.2) Proposition. *The assignment of $\mathfrak{s}_K$ to cones K satisfies the following properties:*

(1) *(Invariance under lattice automorphisms) If K and L are equivalent under an automorphism of $\mathbb{Z}^d$ which preserves the prescribed ordering of the rays of K and L, then*

$$\mathfrak{s}_K(\mathbf{y}) = \mathfrak{s}_L(\mathbf{y}).$$

(2) *(Additivity under subdivision) If $K, K_1, \ldots, K_l$ are d-dimensional rational simple cones whose indicator functions satisfy*

$$[K] = \alpha_1[K_1] + \cdots + \alpha_l[K_l]$$

modulo cones of smaller dimension, then

$$\mathfrak{s}_K(\mathbf{y}) = \alpha_1 \mathfrak{s}_{K_1}(\mathbf{y}KK_1^{-1}) + \cdots + \alpha_l \mathfrak{s}_{K_l}(\mathbf{y}KK_l^{-1}).$$

(3) *(Summation formula) Let Π be the fundamental parallelepiped for K^* and let $u_1, \ldots, u_d$ be the primitive generators for the cone K^*. Then*

$$\mathfrak{s}_K(\mathbf{y}) = \Bigg(\sum_{u \in \Pi \cap \mathbb{Z}^d} e^{-\langle u, \mathbf{y}K \rangle} \Bigg) \prod_{i=1}^{d} \frac{1}{1 - e^{-\langle u_i, v_i \rangle y_i}}.$$

Thus the function $\mathfrak{t}_K$ is regular at the origin and hence defines a power series in the variables $y_1, \ldots, y_n$.

Sketch of Proof. Invariance under lattice automorphisms can be easily checked from the definition of $\mathfrak{s}$; it is also a consequence of the summation formula (Part (3) above), which is clearly invariant under lattice automorphisms.

To check additivity under subdivisions, let $K, K_1, \ldots, K_l$ be as above, and apply Corollary 2.8, which implies that the indicator functions $[K_i^*]$ of the dual cones sum to the indicator function of K^* modulo straight lines. But by Theorem 3.1, Part 2, f vanishes on cones containing straight lines. The desired identity is then a consequence of the additivity formula for f given in Theorem 3.1, Part 1.

The summation formula follows easily from the formula of Example 3.3 and the definitions, bearing in mind that $\langle u_i, \mathbf{y}K \rangle = \langle u_i, v_i \rangle y_i$. Finally, the assertion about the power series $\mathfrak{t}_K(\mathbf{y})$ follows from the summation formula for $\mathfrak{s}_K(\mathbf{y})$ and the fact the function

$$g(z) = \frac{z}{1 - e^{-z}}$$

is regular at $z = 0$. □

One further useful property of $\mathfrak{t}$ is its compatibility along common faces:

(8.3) Lemma. *If cones K and L meet at a common face F, then the power series $\mathfrak{t}_K$ and $\mathfrak{t}_L$ agree when restricted to F; that is, if in $\mathfrak{t}_K$ and $\mathfrak{t}_L$, we set to zero all variables corresponding to rays outside of F, we obtain identical power series in the remaining variables.*

Sketch of Proof. When the variables in $\mathfrak{t}_K$ corresponding to rays not in F are set to zero, one checks that we obtain $\mathfrak{t}_F$, computed with respect to the linear subspace $F+(-F)$. The result is therefore dependent only on F, not on all of K. □

We are now ready to state the Todd class formula. Let Σ be a complete simplicial fan with rays $v_1, \ldots, v_l$. We introduce the Stanley-Reisner ring of the fan

$$A_\Sigma = \frac{\mathbb{Q}[y_1, \ldots, y_l]}{I_\Sigma},$$

where I_Σ denotes the ideal generated by all monomials $y_{i_1} \cdots y_{i_k}$ where $\langle v_{i_1}, \ldots, v_{i_n} \rangle$ is not a cone in the fan Σ.

The power series $\mathfrak{t}_K$ corresponding to the d-dimensional cones of Σ fit together to form a single power series:

(8.4) Definition. Let σ be a complete simplicial fan as above. Since the power series $\mathfrak{t}_K$ where K is a maximal cone of Σ are compatible in the sense of Lemma 8.3, they patch together to form a single power series

$$\mathfrak{t}_\Sigma(y_1, \ldots, y_l)$$

in variables $y_1, \ldots . y_l$ corresponding to the rays $v_1, \ldots, v_l$ of Σ. Since monomials corresponding to cones not in the fan do not appear in $\mathfrak{t}_\Sigma(y_1, \ldots, y_l)$, this power series lives naturally in the completion of the Stanley-Reisner ring A_Σ.

The main theorem of this section states that t_Σ computes the Todd class of the toric variety X_Σ. This theorem was proved in the context of equivariant cohomology by Brion and Vergne [7]. The theorem is also equivalent to the formulas of [55]. In that context, the Todd class formula was a consequence of a local push-forward formula for products of cycles on a toric variety.

(8.5) Theorem. *For any complete simplicial fan Σ, the Todd class of X_Σ is obtained by evaluating $\mathfrak{t}_\Sigma(y_1, \ldots, y_l)$ at $y_i = D_i$, where $D_i \in A^1 X_\Sigma$ is the divisor class corresponding to the ray v_i. That is:*

$$\operatorname{Td} X_\Sigma = \mathfrak{t}_\Sigma(D_1, \ldots D_l).$$

We remark here that if all cones of Σ are unimodular, then the above equation expresses the well-known fact that the Todd classes of a nonsingular toric variety are found by taking the Todd polynomials, introduced in Definition 5.1, in the classes D_i of the torus-invariant divisors. This results from the general fact that the Todd classes of any nonsingular variety can be computed from the Chern classes of the tangent bundle using the Todd polynomials.

In order to determine the Todd class of any simplicial X_Σ of dimension d, only those terms of $\mathfrak{t}_\Sigma$ of degree less than or equal to d need to be computed. This is because any product of degree larger than d represents 0 in the Chow group. With this in mind, it is not hard to see that the above expression for the Todd class may be computed in polynomial time in fixed dimension:

(8.6) Theorem. *For a fixed dimension d, there exists a polynomial time algorithm which, given a complete fan Σ, computes $\mathfrak{t}_\Sigma$ up to degree d. Thus the Todd class Σ can be expressed as a polynomial in the torus-invariant divisors in polynomial time.*

Sketch of proof. The idea is very similar to the proof of Theorem 4.4. Again, the key is the result of [3], stated as Theorem 4.2 above, that a polynomial time algorithm exists to write an arbitrary rational cone as the difference of unimodular cones. To compute the Todd class, one applies this theorem to all d-dimensional cones of Σ, and uses the additivity property of Proposition 8.2 to express the power series $\mathfrak{t}_K$ up to degree d. At this point we are done, since these expressions for $\mathfrak{t}_K$ determine $\mathfrak{t}_\Sigma$. □

The well-known dictionary between polytopes and toric varieties allows us to use this result to obtain a polynomial-time algorithm for computing the number of lattice points in an integral convex polytope. The following algorithm appeared in [55]:

(8.7) Algorithm. Given a simple integral convex polytope P, the following algorithm computes the number of lattice points in P:

(1) Suppose that P is represented as the solution to the finitely many inequalities as follows:

$$P = \{x \in \mathbb{R}^d : \langle v_i, x\rangle \geq \beta_i : i = 1, \ldots l\},$$

where the v_i are the rays of the inner normal fan Σ of P.

(2) Using (8.6), compute the Todd power series $\mathfrak{t}_\Sigma(y_1, \ldots, y_d)$ in degree up to d. Denote this degree d polynomial by T.

(3) Let

$$C = \exp\Big(\sum -\beta_i y_i\Big),$$

and let Q be the degree d part of the product TC.

(4) In the Stanley-Reisner ring A_Σ, let J be the ideal generated by the set

$$\Big\{\sum_{i=1}^{l} \langle v_i, u\rangle y_i : u \in \mathbb{Z}^n\Big\}.$$

Choose any vertex of P, and let $K = \langle v_{i_1}, \ldots, v_{i_d}\rangle$ be the corresponding cone of Σ. Compute normal forms of Q and of $y_{i_1} \cdots y_{i_d}$ with respect to any Gröbner basis for the ideal J. The degree d part of the quotient A_Σ/J is known to be a one-dimensional vector space. Thus these two normal forms may be divided to produce a rational number. The number of lattice points in P is then expressed by:

$$|P \cap \mathbb{Z}^d| = \frac{\mathrm{nf}(Q)}{\mathrm{ind}(K)\mathrm{nf}(y_{i_1} \cdots y_{i_d})}.$$

Sketch of Proof. The fact that the above algorithm yields the number of lattice points in P follows from a well-known application of the Riemann-Roch theorem to the toric variety X_Σ. To the polytope P, there corresponds naturally a line bundle L_P on the toric variety X_Σ. Lattice points in P are in one-to-one correspondence with a basis of global sections of this line bundle. Hence the lattice point question reduces to finding the dimension of the space of sections of L_Σ. The higher cohomology of L_Σ vanishes, and therefore the number of lattice points equals the Euler characteristic $\chi(X_\Sigma, L_P)$. The singular Hirzebruch-Riemann-Roch theorem allows us to compute this Euler characteristic by taking the intersection product of the Todd class of X_Σ, represented by T, with the Chern character of L_P, represented by C. This is expressed in step (3) above. Finally, according to Hirzebruch-Riemann-Roch, one must find the codimension d part of the product $Q = TC$. Step (4) accomplishes this by computing in the ring A_Σ/J, which is a well-known presentation for the Chow ring of X_Σ. The codimension d part of this ring is one-dimensional, consisting of multiples of the class of a point. Furthermore, for any cone

$$K = \langle v_{i_1}, \ldots, v_{i_d} \rangle,$$

as in step (4), the product of the corresponding divisors is given by

$$D_{i_1} \cdots D_{i_d} = \frac{1}{\operatorname{ind}(K)}[\mathrm{pt}].$$

Hence the formula of step (4) expresses Q as a multiple of the class of a point, and hence computes the number of lattice points in P. □

The above algorithm can also be viewed from the point of view of local formulae, in the spirit of Theorem 6.1. To see this, we give the following variation of Algorithm 8.7, which is very convenient in the case that the volumes of the faces of P are known.

Following the above notation, we begin with a simple lattice polytope P, and we compute T as in step (1) above. Using the linear relations of the ideal J, one may easily obtain a *squarefree* expression for T. One then replaces each squarefree monomial $y_{i_1} \cdots y_{i_k}$ occurring in T by

$$\frac{1}{\operatorname{ind}(K)} \operatorname{vol}(F),$$

where K is the k-dimensional cone

$$K = \langle v_{i_1}, \ldots, v_{i_k} \rangle,$$

and F is the corresponding $(d-k)$-dimensional face of P defined by

$$F = \{x \in P : \langle v_i, x \rangle = \beta_i : i = i_1, \ldots, i_k\}.$$

With this replacement, one obtains a rational number, which equals the number of lattice points in P.

We note that this variant of Algorithm 8.7 may be used to compute a given coefficient in the Ehrhart polynomial of P independently of the other coefficients. If we are interested only in the top k Ehrhart coefficients, we must consider only those monomials in T of degree at most k, and only those cones $K \in \Sigma$ of dimension at most k. It follows that for fixed k, the above procedure reduces in polynomial time the computation of the top k Ehrhart coefficients of an integral polytope P to the computation of the volumes of its faces. This provides a partial answer to Problem 6.2.

(8.8) Question. The first polynomial-time algorithm for counting lattice points in polytopes was that of Theorem 4.4 (originally in [3]), which involved no toric varieties. Algorithm 8.7, though based on algebraic geometry, appears quite similar in flavor. In particular, both algorithms are ultimately based on subdividing cones into unimodular cones, and both are linked quite closely to the valuation $\mathfrak{F}$ introduced in Theorem 3.1. In should be noted, however, that the original algorithm is based on subdivisions of the tangent cones of the polytope, whereas Algorithm 8.7 naturally involves subdivisions of the dual cones. These observations motivate several questions: How exactly are these algorithms related? What is the precise nature of the duality involved here? If one were to implement these algorithms, what features of each algorithm would be advantageous?

It should be noted that the power series $\mathfrak{t}$ also plays a key role in the lattice point formula of Brion-Vergne [8, Theorem 2.15], which is a generalization of the formula of Khovanskii-Pukhlikov [33]. In these rather remarkable formulas, the power series $\mathfrak{t}_\Sigma$ is considered as an infinite-order differential operator, called the *Todd differential operator.* A deformed polytope is created, with all facets moved independently, but parallel to the original facets. The volume of this deformed polytope is calculated as a function of the displacements. Applying the Todd differential operator to this function yields the number of lattice points in the polytope. More generally, one obtains an *Euler-Maclaurin formula*, expressing the sum of any polynomial function over the lattice points in a polytope as the Todd operator applied to the integral of the polynomial function over the deformed polytope.

We now state these formulas precisely. As above, assume that P is a simple integral convex polytope. Suppose that

$$P = \{x \in \mathbb{R}^d : \langle v_i, x\rangle \geq \beta_i : i = 1, \ldots l\},$$

where the v_i are the rays of the inner normal fan Σ of P. For $h = (h_1, \ldots, h_l) \in \mathbb{R}^l$, we define the deformed polytope $P(h)$ by:

$$P(h) = \{x \in \mathbb{R}^d : \langle v_i, x\rangle \geq \beta_i - h_i : i = 1, \ldots l\}.$$

We then have:

(8.9) Theorem. *Let P be as above, let Σ be its inner normal fan, and let ϕ be a polynomial function on $\mathbb{R}^d$. Denote by $I_\phi(h)$ the integral of ϕ over the deformed polytope $P(h)$:*

$$I_\phi(h) = \int_{P(h)} \phi(x)dx.$$

Then we have

$$\sum_{m \in P \cap \mathbb{Z}^d} \phi(m) = \mathfrak{t}_\Sigma\left(\frac{\partial}{\partial h_1}, \ldots, \frac{\partial}{\partial h_l}\right) \diamond I_\phi(h).$$

Here the diamond symbol indicates that all derivatives are evaluated at $h = 0$.

In particular,

$$|P \cap \mathbb{Z}^d| = \mathfrak{t}_\Sigma\left(\frac{\partial}{\partial h_1}, \ldots, \frac{\partial}{\partial h_l}\right) \diamond \mathrm{Vol}(P(h)).$$

We have thus far seen several results linking the problem of lattice point enumeration with the valuation $\mathfrak{F}$ and its relatives $\mathfrak{s}$ and $\mathfrak{t}$, which are all efficiently computable. We close this section by showing that these functions are also closely tied with the important formulas of R. Morelli.

In the late 1970's, V. I. Danilov asked for a local expression for the Todd class of a toric variety. Specifically, he asked if there exists a function μ on rational cones such that for any complete fan Σ, the Todd class of X_Σ is given by

$$\operatorname{Td} X_\Sigma = \sum_{K \in \Sigma} \mu(K)[V(K)],$$

where $V(K)$ is the closed subvariety of X_Σ corresponding to the cone $K \in \Sigma$. In 1993, Morelli [50] showed that this assignment was indeed possible. Though this μ is far from unique, Morelli showed that a canonical μ does exist if we allow μ to take values in a certain field of rational functions (rather than simply rational or real values.) Precisely, he showed that for each $k \leq d$, there exists a function μ_k from the set of all k-dimensional rational cones in $\mathbb{R}^d$ to the field $Rat(Gr_{d-k+1}(\mathbb{R}^d))$ of rational functions on the Grassmannian of $d-k+1$ planes in $\mathbb{R}^d$, such that μ_k expresses the Todd class in the sense of Danilov's equation above.

The spirit of this question is quite similar to that of McMullen's Theorem, stated as Theorem 6.1 above. In fact, using the link between Todd classes and lattice point enumeration, Morelli's affirmative answer to the above question immediately settles Theorem 6.1 for integral polytopes, via the relation

$$\phi(K) = \mu(K^*).$$

The following proposition (cf. [55, Proposition 4]) gives a relation between Morelli's $\mu_d(K)$ and the Laurent series $\mathfrak{s}_K$.

(8.10) Proposition. *Let K be a d-dimensional rational simple cone in $\mathbb{Z}^d$, and let $u_1, \dots, u_d$ be the primitive generators of the dual cone K^*. Let $\mathfrak{s}_K^0$ denote the degree 0 part of the Laurent series $\mathfrak{s}_K$. Then $\mathfrak{s}_K^0(u_1, \dots, u_d)$, as a degree 0 rational function on $\mathbb{R}^d$, lives naturally in $Rat(Gr_1(\mathbb{R}^d))$. As such, it coincides with Morelli's $\mu_d(K)$:*

$$\mu_d(K) = \mathfrak{s}_K^0(u_1, \dots, u_d).$$

Sketch of Proof. We can check equality of the above rational functions on unimodular cones easily. By Proposition 8.2, Part 3,

$$\mathfrak{s}_K(u_1, \dots, u_d) = \prod_{i=1}^{d} \frac{1}{1 - e^{-u_i}},$$

the degree 0 part of which equals the d-th Todd polynomial $\operatorname{td}_d(u_1, \dots, u_d)$. This equals $\mu_d(K)$, as discussed in [50] following the statement of Theorem 4. Equality for all simple cones then follows from additivity under subdivisions, satisfied by both Morelli's μ, and $\mathfrak{s}$ (cf. Proposition 8.2, Part 2.) □

It is also possible to relate Morelli's $\mu_k(K)$ for a simple k-dimensional cone K, with $k < d$, to the coefficients of the Laurent series $\mathfrak{s}_K$. A discussion of this connection may be found in [55].

9. Generalized Dedekind Sums and Counting Lattice Points

In Example 1.4, we saw the classical Dedekind sum appear in a formula for the number of lattice points in a certain tetrahedron. In this section we explore the important role that Dedekind sums and their higher-dimensional generalizations play in lattice point formulas. In particular, following Brion-Vergne, we show that the higher-dimensional Dedekind sums introduced by Zagier [63] appear as coefficients in the power series $\mathfrak{t}$ of Section 8. It then follows from the results of Section 8 that these important sums of Zagier are computable in polynomial time when the dimension is fixed.

The classical Dedekind sum first appeared long ago in Dedekind's work on the η-function, and since then has arisen in a variety other contexts. For example, Hirzebruch connected these sums with geometry by showing that they appear naturally in formulas for the signature of certain singular quotient spaces. The 1951 formula of Mordell (Example 1.4) marked the first appearance of Dedekind sums in a formula for lattice points.

We now relate the classical Dedekind sum to the power series $\mathfrak{t}$ introduced in Section 8. This relation, less than a decade old, puts the Mordell formula in a much more general and geometric context.

Let K be a two-dimensional cone in a lattice. All such cones are simplicial, and in fact K is equivalent by a lattice isomorphism to the cone $\langle (0,1),(p,q)\rangle$ in $\mathbb{Z}^2$, with $0 \le p < q$. In this case, the cone K is said to have *type* (p,q). It is easily checked that q (which equals $\operatorname{ind}(K)$) is uniquely determined by K, and p is determined up to multiplicative inverses modulo q. (The cone $\langle (0,1),(p^{-1},q)\rangle$ is equivalent to K by a lattice isomorphism that swaps the rays of K.) This allows us to make the following definition, which associates a Dedekind sum to any two-dimensional cone.

(9.1) Definition. Let K be a two-dimensional cone. If K has type (p,q), then we define the Dedekind sum associated to K, denoted $s(K)$ by:

$$s(K) = s(p,q).$$

This Dedekind sum appears as a coefficient in the degree two part of the power series $\mathfrak{t}_K$:

(9.2) Theorem. *For any two-dimensional cone K, the degree two part of the power series $\mathfrak{t}_K$ is given by:*

$$(\mathfrak{t}_K(y_1,y_2))_{\deg 2} = \frac{1}{12}(y_1^2+y_2^2) + \operatorname{ind}(K)\left(s(K)+\frac{1}{4}\right)y_1y_2.$$

Sketch of Proof. Any one-dimensional cone is unimodular. Thus, by Proposition 8.2, Part 3, the coefficients of y_1^2 and y_2^2 are always $\frac{1}{12}$, the coefficient of z^2 in the expansion of

$$g(z) = \frac{z}{1-e^{-z}}.$$

As for the coefficient of y_1y_2, one computes directly from Proposition 8.2, Part 3. The fundamental parallelepiped Π for K^* has q elements

$$\Pi = \left\{\left(j, \left\lceil \frac{-pj}{q}\right\rceil\right) : j = 0,\ldots,q-1\right\},$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. The desired equality follows by expanding the expression in Proposition 8.2, and using the definition of Dedekind sum given in Example 1.4.

Alternatively, one could use the (dual) approach of the proof of Theorem 9.5 (see below.)

□

The above theorem, together with the ideas of Algorithm 8.7, allow one to express the degree $d-2$ coefficient of the Ehrhart polynomial of an arbitrary d-dimensional lattice polytope. This expression involves one Dedekind sum for each face of codimension two. For each such face F, the dual to the tangent cone at F is a two-dimensional cone K, and the corresponding Dedekind sum $s(K)$ contributes to the degree $d-2$ term in the Ehrhart polynomial.

These ideas are especially valuable for polytopes of dimensions two and three. In dimension three, the only mysterious part of the Ehrhart polynomial is its linear term. Thus, we see that the entire Ehrhart polynomial may be easily computed in terms of the Dedekind sums corresponding to the edges (1-dimensional faces) of the polytope.

For $d=2$, the degree $d-2$ part of the Ehrhart polynomial simply equals 1 for any polytope. Thus it would seem that a Dedekind sum expression for this coefficient would be useless. However, when we equate this Dedekind sum expression to 1, we obtain a reciprocity relation for Dedekind sums! In this way, any lattice polygon gives a relation among the Dedekind sums corresponding to its angles (see [53, Theorem 8].) One easily finds lattice triangles that demonstrate Dedekind's reciprocity relation, Rademacher's three-term law, as well as new relations for the classical Dedekind sum.

We now turn our attention to the higher dimensional Dedekind sums introduced by Zagier. Motivated by topological considerations, Zagier made the following definition:

(9.3) Definition. Let q be a positive integer, and let $p_1, \ldots, p_n$ be integers prime to q, with n even. The higher dimensional Dedekind sum $d(q; p_1, \ldots, p_n)$ is defined by:

$$d(q;p_1,\ldots,p_n) = (-1)^{\frac{n}{2}} \sum_{k=1}^{q-1} \cot\frac{\pi k p_1}{q} \cdots \cot\frac{\pi k p_n}{q}.$$

Note that if n is odd, the above sum clearly vanishes. It should be noted that one can also define the higher dimensional Dedekind sum in terms of the hyperbolic cotangent:

$$d(q;p_1,\ldots,p_n) = \sum_{k=1}^{q-1} \coth\frac{\pi k p_1 i}{q} \cdots \coth\frac{\pi k p_n i}{q}.$$

The name "higher dimensional Dedekind sum" is justified in part by the equality

$$d(q;p_1,p_2) = -\frac{2}{3} s(p_1 p_2^{-1}, q).$$

That is, when $n=2$, the higher dimensional Dedekind sum reduces to a classical Dedekind sum.

Zagier showed that his sums satisfy a reciprocity relation which generalizes the reciprocity formula for the classical Dedekind sum:

(9.4) Theorem. *Let $a_0, \dots a_n$ be coprime integers, with n even. Then*

$$\sum_{j=0}^{n} \frac{1}{a_j} d(a_j; a_0, \dots, \hat{a_j}, \dots, a_n) = 1 - \frac{L_n(a_0, \dots, a_n)}{a_0 \cdots a_n}.$$

Here $L_n(a_0, \dots, a_n)$ is the nth L-polynomial, defined as the coefficient of t^n in the power series expansion of

$$\prod_{j=0}^{n} \frac{a_j t}{\tanh a_j t}.$$

Like the classical reciprocity law, this theorem may be proved using the connection with Todd classes: see the remark following Theorem 9.5.

While Theorem 9.4 is easily seen to be a generalization of Dedekind's classical reciprocity law, there is an important difference. As we have seen, the classical reciprocity law allows for efficient computation of the classical Dedekind sum. However, in the case of the higher dimensional sums, it is not at all clear how to use Theorem 9.4 to compute these sums efficiently. In fact, as Zagier points out, it is not even clear if this reciprocity law (together with obvious symmetry and periodicity properties) *characterizes* these sums. Below (Theorem 9.6), we show that these higher dimensional sums can be computed in polynomial time when the dimension is fixed.

We next show that higher dimensional Dedekind sums appear as coefficients in the power series $\mathfrak{t}$ introduced in Section 8. The following theorem is due to Brion and Vergne. The cotangent formula below also appears in the important work of Diaz-Robins [13], which is discussed briefly at the end of this section.

(9.5) Theorem. *Let n be even, and let K be a n-dimensional simplicial cone in $\mathbb{Z}^n$. Suppose that all faces of K of dimension $n-1$ are unimodular. This means that after a change of basis, the primitive lattice points on the rays of K may be chosen as:*

$$\begin{aligned} v_1 &= (1, 0, \dots, 0) \\ v_2 &= (0, 1, \dots, 0) \\ &\cdot \\ &\cdot \\ v_n &= (p_1, \dots, p_{n-1}, q), \end{aligned}$$

with $q = \operatorname{ind}(K) > 0$. We then have the following expression for the coefficient of $y_1 \cdots y_d$ in the power series $\mathfrak{t}(y_1, \dots y_n)$:

$$\frac{1}{2^n q} \sum_{k=1}^{q} \left(1 + \coth\left(\frac{\pi k p_1 i}{q}\right)\right) \cdots \left(1 + \coth\left(\frac{\pi k p_{n-1} i}{q}\right)\right) \left(1 + \coth\left(\frac{\pi k i}{q}\right)\right)$$

Sketch of Proof. Here it is useful to have a dual expression for the power series $\mathfrak{t}_K$, which is formally equivalent to Part 3 of Proposition 8.2. Let $u_1, \dots, u_n$ denote the primitive elements of the dual lattice satisfying $\langle u_i, v_j \rangle = 0$ for $i \neq j$, and for $v \in \mathbb{Z}^n$, define

$$a_j(v) = \exp\left\{2\pi i \frac{\langle u_j, v \rangle}{\langle u_j, v_j \rangle}\right\}.$$

Let Π_K denote the fundamental parallelepiped for K. (Note this differs from Π above, which is the parallelepiped for the dual cone K^*.) We then have the formula (cf. [7])

$$\mathfrak{s}_K(\mathbf{y}) = \frac{1}{\operatorname{ind}(K)} \sum_{v\in\Pi_K} \prod_{j=1}^{n} \frac{1}{1-a_j(v)e^{-y_i}}.$$

The desired coefficient of $y_1 \cdots y_n$ is then easily computed, keeping in mind that

$$\frac{1}{1-e^{-z}} = 1 + \coth\left(\frac{z}{2}\right).$$

□

Notice that by multiplying out the product appearing in Theorem 9.5, one obtains an expression for this coefficient as a sum of higher dimensional Dedekind sums. The "leading term" is the n-dimensional Dedekind sum $d(q; p_1, \ldots, p_{n-1}, 1)$.

We also remark here that the above formula can be used to give a simple geometric proof of Zagier's reciprocity formula (Theorem 9.4). In $\mathbb{Z}^n$, take the unimodular cone K spanned by the standard unit vectors. Subdivide K into n cones by introducing the ray through $(b_1, \ldots b_n)$, where the b_i's are pairwise coprime positive integers. Applying the additivity formula for $\mathfrak{t}$ (Proposition 8.2) to this subdivision yields a relation on the coefficients in the power series associated to these cones. Since Theorem 9.5 identifies certain of these coefficients as Dedekind sums, we get a relation among Dedekind sums. This relation is easily seen to yield Zagier's higher reciprocity law.

We now show that the results of Section 8 provide for efficient computation of the higher dimensional Dedekind sum, a result not obvious from the reciprocity law (Theorem 9.4).

(9.6) Theorem. *For fixed dimension n, there is a polynomial time algorithm which, given integers $q > 0$ and $p_1, \ldots p_n$ relatively prime to q, computes the higher dimensional Dedekind sum $d(q; p_1, \ldots, p_n)$*

Sketch of Proof. First notice that we can reduce to the case in which $p_n = 1$. This is because of the easily verified identity:

$$d(q; p_1, \ldots, p_n) = d(q; p_n^{-1} p_1, \ldots, p_n^{-1} p_{n-1}, 1),$$

where p_n^{-1} is a multiplicative inverse of p_n modulo q.

Applying Theorem 8.6, we see that the product of Theorem 9.5 is computable in polynomial time for fixed dimension n. This product, when expanded, can be expressed as the sum of 2^n higher dimensional Dedekind sums (many of which are zero). Of these sums, all but the leading term, $d(q; p_1, \ldots, p_{n-1}, 1)$ are Dedekind sums of dimension less than d. By induction, these may be computed in polynomial time. Thus, $d(q; p_1, \ldots, p_{n-1}, 1)$ itself may be computed in polynomial time. □

Theorem 9.5 links a certain coefficient in the power series $\mathfrak{t}_K$ with the higher dimensional Dedekind sum. This Dedekind sum appears in the very special case when all facets of the cone K are unimodular. For general cones K, the general coefficient in $\mathfrak{t}_K$ represents a significant further generalization of Zagier's sums. These coefficients also admit cotangent formulas, using the ideas of the proof of Theorem 9.5.

In fact, these more general cotangent sums can be seen in the remarkable lattice point formula of Diaz-Robins [13]. This very explicit formula expresses the number of lattice points in an arbitrary simplex in terms of cotangent sums. The sums that appear are easily seen to match the cotangent sums that appear as coefficients in the power series $\mathfrak{t}$. The cotangent formulas Diaz and Robins give are expressed in an explicit and appealing form. However, as written they do not appear to be efficiently computable. This is because the size of these sums may be as large as the indices of the tangent cones to the polytope. As discussed in Section 4, these indices are not bounded by a polynomial in the input size of the cone. Nevertheless, because these sums may be expressed in terms of the coefficients of $\mathfrak{t}$, they are in fact computable in polynomial time in fixed dimension, by Theorem 8.6. The forthcoming paper [56] will explore details of this connection, as well as reciprocity relations satisfied by these interesting cotangent sums.

10. What is the Complexity of the Presburger Arithmetic?

In Section 4, we saw that the generating function for the set of integer points in a polyhedron P,

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

has a "short" (that is, polynomial in the input size of P) representation as a rational function. Let us switch gears and let us consider sets $S \subset \mathbb{Z}^d$ that are *given* by their rational generating function

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

where $\alpha_i \in \mathbb{Q}$ and $p_i; a_{i1}, \ldots, a_{is} \in \mathbb{Z}^d$, which can be expanded into the Laurent series

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

There is an ambiguity here since $f(S;\mathbf{x})$ may have a number of different Laurent expansions. To fix this, we assume that the set S is finite: from the computational complexity point of view, there is not a big difference between infinite and very large finite sets.

We will be interested in the sets S that can be encoded by a "short" rational function $f(S;\mathbf{x})$. In particular, we will assume that both the number d of variables and the number k of binomials $(1-\mathbf{x}^{a_{ij}})$ in the denominator of each fraction are fixed. Our main thesis is that the set $S \subset \mathbb{Z}^d$ is computationally tractable if and only if it can be encoded by a short rational function.

(10.1) Examples. Suppose that a and b are coprime positive integers. Let $S = \{\alpha a + \beta b : \ \alpha, \beta \in \mathbb{Z}_+\}$ be the set of all non-negative integer combinations of a and b. Then, for $|x| < 1$ we have

$$f(S;x) = \sum_{m \in S} x^m = \frac{1-x^{ab}}{(1-x^a)(1-x^b)}.$$

The set S contains all sufficiently large positive integers, so the initial interval of S is the main interest. We can write

$$f(S;x) = p(x) + \frac{x^N}{1-x}$$

for some polynomial $p(x)$ and some positive integer N. The polynomial $p(x)$ encodes in a compact way information regarding the initial interval of S.

Suppose that a, b, and c are coprime positive integers. Let $S = \{\alpha a + \beta b + \gamma c : \alpha, \beta, \gamma \in \mathbb{Z}_+\}$ be the set of all non-negative integer combinations of a, b, and c. Then for $|x| < 1$, we have

$$f(S;x) = \frac{1 - x^{n_1} - x^{n_2} - x^{n_3} + x^{n_4} + x^{n_5}}{(1-x^a)(1-x^b)(1-x^c)},$$

for certain positive integers n_1, n_2, n_3, n_4, and n_5. This interesting fact was discovered, apparently, by G. Denham [12]. This result follows also from a general result of [51]. The paper [62] contains an elementary proof that the number of terms in the numerator is at most 12, which is twice as many as the sharp bound.

For example, if $a = 23$, $b = 29$, and $c = 44$, then

$$f(S;x) = \frac{1 - x^{161} - x^{203} - x^{220} + x^{249} + x^{335}}{(1-x^{23})(1-x^{29})(1-x^{44})}.$$

As in the previous example, the set S contains all sufficiently large positive integers, so the initial interval of S is the interesting part. It can be encoded by a short polynomial.

QUESTION. What information regarding S can be extracted from $f(S;\mathbf{x})$ and what operations on sets given by their generating functions $f(S;\mathbf{x})$ can be carried out efficiently?

For example, by substituting $\mathbf{x} = (1, \ldots, 1)$, one can get the number of points in S (since $\mathbf{x} = (1, \ldots, 1)$ is the pole of every fraction, one may need to compute an appropriate residue, as in Section 5).

Our next result is that one can efficiently perform Boolean operations on sets given by their functions $f(S;\mathbf{x})$. In fact, we prove a more general statement: the Hadamard product of short rational functions can be computed in polynomial time.

(10.2) Theorem. *Suppose that $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are rational functions in d complex variables $\mathbf{x} = (x_1, \ldots, x_d)$:*

$$f_1(\mathbf{x}) = \sum_{i \in I_1} \alpha_i \frac{\mathbf{x}^{p_i}}{(1-\mathbf{x}^{a_{i1}}) \cdots (1-\mathbf{x}^{a_{ik}})}, \quad f_2(\mathbf{x}) = \sum_{i \in I_2} \beta_i \frac{\mathbf{x}^{q_i}}{(1-\mathbf{x}^{b_{i1}}) \cdots (1-\mathbf{x}^{b_{ik}})},$$

where $\alpha_i, \beta_i \in \mathbb{Q}$ and $p_i, a_{i1}, \ldots, a_{ik}; q_i, b_{i1}, \ldots, b_{ik} \in \mathbb{Z}^d$.

Suppose further that

$$f_1(\mathbf{x}) = \sum_{m \in \mathbb{Z}^d} \gamma_m \mathbf{x}^m \quad \textit{and} \quad f_2(\mathbf{x}) = \sum_{m \in \mathbb{Z}^d} \delta_m \mathbf{x}^m$$

are the Laurent expansions of f_1 *and* f_2*, respectively, for* $\mathbf{x}$ *in a non-empty open set* $U \subset \mathbb{C}^d$*, such that* $|\mathbf{x}^{a_{ij}}| < 1$ *and* $|\mathbf{x}^{b_{ij}}| < 1$ *for all* i, j *and all* $\mathbf{x} \in U$.

Then there exists a rational function $f(\mathbf{x}) = f_1(\mathbf{x}) \star f_2(\mathbf{x})$*, called the Hadamard product of* f_1 *and* f_2*,*

$$f(\mathbf{x}) = \sum_{i \in I} \epsilon_i \frac{\mathbf{x}^{r_i}}{(1 - \mathbf{x}^{c_{i1}}) \cdots (1 - \mathbf{x}^{c_{i(2k)}})}$$

with the Laurent expansion

$$f(\mathbf{x}) = \sum_{m \in \mathbb{Z}^d} (\gamma_m \delta_m) \mathbf{x}^m$$

for $\mathbf{x} \in U$*. Moreover, for fixed* d *and* k *there is a polynomial time algorithm for computing* $f(\mathbf{x})$ *from* $f_1(\mathbf{x})$ *and* $f_2(\mathbf{x})$*.*

Sketch of Proof. Since the Hadamard product is a bilinear operation, it suffices to compute $f_1(\mathbf{x}) \star f_2(\mathbf{x})$ in the particular case of the simple fractions:

$$f_1(\mathbf{x}) = \frac{\mathbf{x}^p}{(1 - \mathbf{x}^{a_1}) \cdots (1 - \mathbf{x}^{a_k})} \quad \text{and} \quad f_2(\mathbf{x}) = \frac{\mathbf{x}^q}{(1 - \mathbf{x}^{b_1}) \cdots (1 - \mathbf{x}^{b_k})}.$$

Expanding both fractions as multiple geometric series, we get

$$f_1(\mathbf{x}) = \sum_{\mu_1, \ldots, \mu_k \geq 0} \mathbf{x}^{p + \mu_1 a_1 + \ldots + \mu_k a_k} \quad \text{and} \quad f_2(\mathbf{x}) = \sum_{\nu_1, \ldots, \nu_k \geq 0} \mathbf{x}^{q + \nu_1 b_1 + \ldots + \nu_k b_k}.$$

In the space $\mathbb{R}^{2k} = \{(\xi_1, \ldots, \xi_k; \eta_1, \ldots, \eta_k)\}$, let P be the rational polyhedron defined by the equations

$$p + \xi_1 a_1 + \ldots + \xi_k a_k = q + \eta_1 b_1 + \ldots + \eta_k b_k$$

and the inequalities

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

Let $\mathbb{Z}^{2k} \subset \mathbb{R}^{2k}$ be the integer lattice $\{(\mu_1, \ldots, \mu_k; \nu_1, \ldots, \nu_k) : \mu_i, \nu_j \in \mathbb{Z}\}$.

Since

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

and the Hadamard product is bilinear, we can write the series for $f(\mathbf{x}) = f_1(\mathbf{x}) \star f_2(\mathbf{x})$ as a sum over the set of integer points in the polyhedron P:

$$f(\mathbf{x}) = \sum_{P \cap \mathbb{Z}^{2k}} \mathbf{x}^{p + \mu_1 a_1 + \ldots + \mu_k a_k} = \mathbf{x}^p \sum_{P \cap \mathbb{Z}^{2k}} \mathbf{x}^{\mu_1 a_1 + \ldots + \mu_k a_k}.$$

By Theorem 4.4, the series

$$\sum_{(m,n) \in P \cap \mathbb{Z}^{2k}} \mathbf{y}^m \mathbf{z}^n,$$

where $m = (\mu_1, \ldots, \mu_k)$, $n = (\nu_1, \ldots, \nu_k)$ and $\mathbf{y}, \mathbf{z} \in \mathbb{C}^k$ can be computed in polynomial time as a rational function $F(\mathbf{y}, \mathbf{z})$. The function $f(\mathbf{x})$ is obtained by specializing $\mathbf{x}^p F(\mathbf{y}, \mathbf{z})$ for $\mathbf{z} = (1, \ldots, 1)$ and $y_i = \mathbf{x}^{a_i}$. One may have to resolve singularities as in Section 5. □

Remark. If the functions f_1 and f_2 are sufficiently generic, that is, if the vectors $a_{i1}, \ldots, a_{ik}; b_{i1}, \ldots, b_{ik}$ span $\mathbb{R}^d$, we will have dim $P = 2k - d$ and

$$f(\mathbf{x}) = \sum_{i \in I} \epsilon_i \frac{\mathbf{x}^{r_i}}{(1 - \mathbf{x}^{c_{i1}}) \cdots (1 - \mathbf{x}^{c_{i(2k-d)}})}.$$

(10.3) Corollary. *Let us fix d and k. There exists a polynomial time algorithm which, for any finite sets $S_1 \subset \mathbb{Z}^d$ and $S_2 \subset \mathbb{Z}^d$, given by their generating functions*

$$f(S_1; \mathbf{x}) = \sum_{i \in I_1} \alpha_i \frac{\mathbf{x}^{p_i}}{(1-\mathbf{x}^{a_{i1}}) \cdots (1-\mathbf{x}^{a_{ik}})}$$

and

$$f(S_2; \mathbf{x}) = \sum_{i \in I_2} \beta_i \frac{\mathbf{x}^{q_i}}{(1-\mathbf{x}^{b_{i1}}) \cdots (1-\mathbf{x}^{b_{ik}})},$$

where $\alpha_i, \beta_i \in \mathbb{Q}$ and $p_i, a_{i1}, \ldots, a_{ik}; q_i, b_{i1}, \ldots, b_{ik} \in \mathbb{Z}^d$, computes the generating functions $f(S_1 \cup S_2; \mathbf{x})$, $f(S_1 \cap S_2; \mathbf{x})$, and $f(S_1 \setminus S_2; \mathbf{x})$.

Proof. Let us choose a generic vector $c \in \mathbb{R}^d$, such that $\langle c, a_{ij} \rangle \neq 0$ and $\langle c, b_{ij} \rangle \neq 0$ for all vectors a_{ij} and b_{ij} By multiplying, if necessary, the denominator and the numerator of each fraction by an appropriate monomial, we can always assume that $\langle c, a_{ij} \rangle < 0$ and $\langle c, b_{ij} \rangle < 0$ for all a_{ij} and b_{ij}. Then the set

$$U = \Big\{ \mathbf{x} \in \mathbb{C}^d : |\mathbf{x}^{a_{ij}}| < 1 : \ i \in I_1, \ j = 1, \ldots, k \quad \text{and}$$

$$|\mathbf{x}^{b_{ij}}| < 1 : \ i \in I_2, \ j = 1, \ldots, k \Big\}$$

is a non-empty open set in $\mathbb{C}^d$, and for every $\mathbf{x} \in U$, there are Laurent expansions:

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

The corollary now follows from Theorem 10.2, since $f(S_1 \cap S_2; \mathbf{x}) = f(S_1; \mathbf{x}) \star f(S_2; \mathbf{x})$, $f(S_1 \cup S_2; \mathbf{x}) = f(S_1; \mathbf{x}) + f(S_2; \mathbf{x}) - f(S_1 \cap S_2; \mathbf{x})$, and $f(S_1 \setminus S_2; \mathbf{x}) = f(S_1; \mathbf{x}) - f(S_1 \cap S_2; \mathbf{x})$. □

The most intriguing question is whether the *projection* of a set with a short generating function is a set with a short generating function.

QUESTION. Let $\pi : \mathbb{Z}^d \longrightarrow \mathbb{Z}^{d-1}$ be the projection

$$\pi(\xi_1, \ldots, \xi_d) = (\xi_1, \ldots, \xi_{d-1}).$$

Let $S \subset \mathbb{Z}^d$ be a set given by its rational generating function

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

Assuming that d and k are fixed, is there a polynomial time algorithm for computing $f\big(\pi(S); \mathbf{y}\big)$, $\mathbf{y} \in \mathbb{C}^{d-1}$, from $f(S; \mathbf{x})$?

An affirmative answer to this question would lead to the solution of a long-standing open problem about the complexity of the Presburger arithmetic. Namely, let us consider a formula with quantifiers and Boolean operations which involves integer variables with the usual order <, additive operations "+" and "−", and multiplication onto integer constants, but not variables. One can come with an algorithm to verify the truth/falsity of such a formula, but as M. Fischer and M. Rabin proved, any such algorithm will have a double exponential complexity [21]. But what if the number of variables and Boolean operations is fixed and not a part of the input? It has long been suspected that then the problem admits a polynomial time algorithm.

Indeed, suppose that the answer to the question is "YES". If we start with the set of integer points in a given rational polytope $P \subset \mathbb{R}^d$ (whose generating function by Theorem 4.4 can be computed in polynomial time) and apply a sequence of projections and Boolean operations (see Corollary 10.3), we get a set of points described by a polynomially computable rational function, provided the dimension d and the number of Boolean operations is fixed. Generally, this way we get a set of points described by a formula of the *Presburger arithmetic*

$$x \in \mathbb{Z}^d : \quad \exists \xi_1 \in \mathbb{Z}\ \forall \xi_2 \in \mathbb{Z}\ \exists \xi_3 \in \mathbb{Z} \ldots \forall \xi_k \in \mathbb{Z} : \quad F(x, \xi_1, \ldots, \xi_k),$$

where F is a quantifier-free formula involving linear inequalities with constant rational coefficients and Boolean operations. In other words, if projection preserves "short" rational functions, the set of points described by a formula of Presburger arithmetic has a generating function whose size is bounded by a polynomial in the size of the coefficients of the formula, provided the number of variables and the number of Boolean operations is fixed. In particular, there would be a polynomial time algorithm for testing the truth/falsity of a formula of Presburger arithmetic, provided the number of variables and the number of Boolean operations is fixed. At present, a polynomial time algorithm for this problem is known if there is at most one quantifier alteration [27]. For example, the following question, known as the Frobenius problem, can be posed as a problem with one quantifier alteration: given coprime positive integers $a_1, \ldots, a_d$ and an integer N, is it true that any integer number $n \geq N$ can be represented as a non-negative integer combination of $a_1, \ldots, a_d$? For a fixed d, a polynomial time algorithm for this problem was constructed by R. Kannan [28].

Sometimes it is easy to see that the projection indeed has a short generating function.

(10.4) Example. Let $P \subset \mathbb{Z}^d$ be a rational polytope, and let $S = P \cap \mathbb{Z}^d$ be the set of integer points in P. Then projection $\pi(S) \subset \mathbb{Z}^{d-1}$ has a generating function whose complexity is polynomial in the input size of P. (Note that $\pi(S)$ is *not* the set of integer points in a rational polytope.) Let $a = (0, \ldots, 0, 1) \in \mathbb{R}^d$ be a vector and let $Q = P \setminus (P + a)$. For $m \in \pi(S)$, the preimage $\pi^{-1}(m) \cap S$ is the set of integer points in the interval $\pi^{-1}(m) \cap P$. It then follows that $m \in \pi(S)$ if and only if there is exactly one integer point $n \in Q$ such that $\pi(n) = m$. Hence, $f(\pi(S); \mathbf{y}) : \mathbf{y} \in \mathbb{C}^{d-1}$ is the specialization of the generating function $f(Q \cap \mathbb{Z}^d; \mathbf{x}) : \mathbf{x} = (\mathbf{y}, z) \in \mathbb{C}^d$ when $z = 1$. Using Theorem 4.4, we conclude that $f(Q \cap \mathbb{Z}^d; \mathbf{x})$ can be computed in polynomial time. Specialization at $z = 1$ may require resolution of singularities as in Section 5 (if the expression for $f(Q \cap \mathbb{Z}^d; \mathbf{x})$ contains fractions with a monomial $(1 - z^k)$ in the denominator). Hence we get a polynomial time algorithm for computing $f(\pi(S); \mathbf{y})$.

(10.5) Definition. Let us fix the decomposition $\mathbb{Z}^d = \mathbb{Z}^l \oplus \mathbb{Z}^{d-l}$, and let $\pi : \mathbb{Z}^d \longrightarrow \mathbb{Z}^{d-l}$ be the projection on the second summand.

For a set $S \subset \mathbb{Z}^d$ and a point $m \in \mathbb{Z}^{d-l}$, we define the *fiber* $\pi^{-1}(m) \subset \mathbb{Z}^l$ as

$$\pi^{-1}(m) = \Big\{ n \in \mathbb{Z}^l : (n, m) \in S \Big\}.$$

We have $\pi^{-1}(m) \subset \mathbb{Z}^l$, and we can consider the generating function $f(\pi^{-1}(m); \mathbf{z})$ for $\mathbf{z} \in \mathbb{C}^l$.

One can prove that if $S \subset \mathbb{Z}^d$ is described by a short rational function, then the fibers $\pi^{-1}(m)$ are described by a "consistent" system of short rational functions as $m \in \mathbb{Z}^{d-l}$ changes.

We use the term *open polyhedron* to mean the relative interior of a polyhedron.

(10.6) Theorem. *Let $S \subset \mathbb{Z}^d$ be a finite set with the generating function*

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

Then the space $\mathbb{R}^l$ can be represented as a disjoint union of open polyhedra Q_j, and for every Q_j, there exist vectors $b_{i1}, \ldots, b_{ik} \subset \mathbb{Z}^l : i \in I_j$, such that for every $m \in Q_j \cap \mathbb{Z}^l$,

$$f\left(\pi^{-1}(m); \mathbf{z}\right) = \sum_{i \in I_j} \beta_i \frac{\mathbf{z}^{q_i}}{(1 - \mathbf{z}^{b_{i1}}) \cdots (1 - \mathbf{z}^{b_{ik}})},$$

where $\beta_i = \beta_i(m)$ are rational numbers and $q_i = q_i(m) \in \mathbb{Z}^l$. In other words, as long as m stays within Q_j, the denominators of the fractions do not change. If d and k are fixed, there is a polynomial time algorithm which, given $f(S; \mathbf{x})$, computes the decomposition of $\mathbb{R}^{d-l}$ into pieces Q_j and the vectors b_{ij}.

Sketch of Proof. As in the proof of Corollary 10.3, we may assume that $|\mathbf{x}^{a_{ij}}| < 1$ for all $\mathbf{x}$ in some open set $U \subset \mathbb{C}^d$. We may write

$$f(S; \mathbf{x}) = \sum_{i \in I} \alpha_i \sum_{\mu_1, \ldots, \mu_k \geq 0} \mathbf{x}^{p_i + \mu_1 a_{i1} + \ldots + \mu_k a_{ik}}.$$

Let $A_i(m) \subset \mathbb{R}^k$ be the affine subspace consisting of the points $(\mu_1, \ldots, \mu_k)$ such that $\pi(p_i + \mu_1 a_{i1} + \ldots + \mu_k a_{ik}) = m$. Splitting $\mathbf{x} = (\mathbf{y}, \mathbf{z})$, where $\mathbf{y} \in \mathbb{C}^{d-l}$ and $\mathbf{z} \in \mathbb{C}^l$, we may write the generating function $f(\pi^{-1}(m), \mathbf{z})$ of the fiber as a specialization of

$$\sum_{i \in I} \alpha_i \sum_{(\mu_1, \ldots, \mu_k) \in \mathbb{Z}^k_+ \cap A_i(m)} \mathbf{x}^{p_i + \mu_1 b_{i1} + \ldots + \mu_k b_{ik}},$$

at $\mathbf{y} = (1, \ldots, 1)$. Now for every $i \in I$, the set $\mathbb{Z}^k_+ \cap A_i(m)$ is the set of integer points in a polyhedron whose facets are moved parallel to themselves as $m \in \mathbb{R}^{d-l}$ changes. Let Q_j be a partition of $\mathbb{R}^{d-l}$ such that the combinatorial type of every polytope $\mathbb{R}^k_+ \cap A_i(m)$ stays the same as long as m changes within Q_j. The proof now follows from Theorem 4.4. □

Remark. From Theorem 4.4, we can deduce that $\beta_i(m)$ and $q_i(m)$ can be expressed by a polynomial size formula, involving arithmetic operations, the "floor" function $\lfloor \cdot \rfloor$, and Boolean functions.

An important feature of Example 10.4 is that the generating function of each fiber is *very short.* Since every fiber is an interval, we have $f(\pi^{-1}(m);z) = (z^a - z^b)/(1-z)$. This observation can be generalized: one can show that if the function $f(\pi^{-1}(m);\mathbf{z})$ contains only a fixed number of terms, then the complexity of the generating function $f(\pi(S);\mathbf{y})$ is bounded by a polynomial in the input size of the function $f(S;\mathbf{x})$.

(10.7) Example. Let $b_1, \ldots, b_d$ be coprime positive integers, let $B = b_1, \ldots, b_d$ be their product and let $a_i = B/b_i$. Let $S = \mathbb{Z}^d_+$, and let $\pi : \mathbb{Z}^d \longrightarrow \mathbb{Z}$ be the projection: $\pi(\alpha_1, \ldots, \alpha_d) = \alpha_1 a_1 + \ldots + \alpha_d a_d$. Then $\pi(S) \subset \mathbb{Z}$ is the set of all non-negative integer combinations of a_i. The set S can be represented as a projection of the non-negative integer orthant $\pi : \mathbb{Z}^d \longrightarrow \mathbb{Z}$.

One can show that the fiber $\pi^{-1}(m)$ is the set of integer points in a totally unimodular simplex (all tangent cones are rational shifts of unimodular cones). Here the generating function of each fiber contains d terms; that is, it has a fixed number of terms provided d is fixed.

It turns out that the set $\pi(S)$ has a short generating function:

$$f(\pi(S);x) = \frac{(1-x^B)^{d-1}}{(1-x^{a_1})\cdots(1-x^{a_d})}.$$

Acknowledgment

The first author was partially supported by the Alfred P. Sloan Research Fellowship, by NSF grants DMS 9501129 and DMS 9734138, and by the Mathematical Sciences Research Institute, Berkeley, CA through NSF grant DMS 9022140. The second author was partially supported by NSF grant DMS 9508972. We wish to thank Michel Brion, Stavros Garoufalidis, Sinai Robins, and Bernd Sturmfels for useful conversations.

References

1. A. Barvinok, Computing the Ehrhart polynomial of a convex lattice polytope, *preprint* TRITA-MAT-1992-0036(Oct 1992), Royal Institute of Technology, Stockholm.
2. A. Barvinok, Computing the volume, counting integral points, and exponential sums, *Discrete & Computational Geometry*, **10**(1993), 123–141.
3. A. Barvinok, A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed, *Mathematics of Operations Research*, **19**(1994), 769–779.
4. A. Barvinok, Computing the Ehrhart polynomial of a convex lattice polytope, *Discrete & Computational Geometry* **12**(1994), 35–48.
5. U. Betke and M. Kneser, Zerlegungen und Bewertungen von Gitterpolytopen. (German) [Decompositions and valuations of lattice polytopes], *Journal für die Reine und Angewandte Mathematik*, **358**(1985), 202–208.
6. M. Brion, Points entiers dans les polyèdres convexes, *Annales Scientifiques de l'Ecole Normale Supérieure (Ser. 4)* , **21**(1988), 653–663.
7. M. Brion and M. Vergne, An equivariant Riemann-Roch theorem for simplicial toric varieties, *Journal für die Reine und Angewandte Mathematik*, **482**(1997), 67–92.

8. M. Brion and M. Vergne, Lattice points in simple polytopes, *Journal of the American Mathematical Society*, **10**(1997), 371–392 .
9. M. Brion and M. Vergne, Residue formulae, vector partition functions and lattice points in rational polytopes, *Journal of the American Mathematical Society*, **10**(1997), 797–833.
10. S. Cappell and J. Shaneson, Genera of algebraic varieties and counting lattice points, *Bulletin of the American Mathematical Society*, **30**(1994), 62–69.
11. W. Cook, M. Hartmann, R. Kannan, and C. McDiarmid, On integer points in polyhedra, *Combinatorica*, **12**(1992), 27–37.
12. G. Denham, The Hilbert series of a certain module, *unpublished manuscript*, 1996.
13. R. Diaz and S. Robins, The Ehrhart polynomial of a lattice polytope, *Annals of Mathematics*, **145**(1997), 503–518.
14. R. Diaz and S. Robins, Solid angles and a Fourier decomposition of lattice polytopes, *to appear.*
15. M. Dyer, On counting lattice points in polyhedra, *SIAM J. Comput.*, **20**(1991), 695–707.
16. M. Dyer, A. Frieze, and R. Kannan, A random polynomial-time algorithm for approximating the volume of convex bodies, *Journal of the Association for Computing Machinery*, **38** (1991), no. 1, 1–17.
17. M. Dyer and R. Kannan, On Barvinok's algorithm for counting lattice points in fixed dimension, *Mathematics of Operations Research*, **22**(1997), 545–549.
18. M. Dyer, A. Frieze, R. Kannan, A. Kapoor, L. Perkovic, and U. Vazirani, A mildly exponential time algorithm for approximating the number of solutions to a multidimensional knapsack problem, *Combinatorics, Probability and Computing*, **2** (1993), 271–284.
19. M. Dyer, R. Kannan, and J. Mount, Sampling contingency tables, *Random Structures & Algorithms*, **10** (1997), 487–506.
20. E. Ehrhart, *Polynômes Arithmétiques et Méthode des Polyèdres en Combinatoire*, International Series of Numerical Mathematics, Vol. 35. Birkhäuser Verlag, Basel-Stuttgart, 1977.
21. M.J. Fischer and M.O. Rabin, Super-exponential complexity of Presburger arithmetic, in: *Complexity of computation* (Proceedings of the SIAM-AMS Symposium, New York, 1973), pp. 27–41. *SIAM-AMS Proc.*, Vol. VII, American Mathematical Society, Providence, R.I., 1974.
22. W. Fulton, *Introduction to Toric Varieties*, Annals of Mathematics Studies, 131. The William H. Roever Lectures in Geometry. Princeton University Press, Princeton, NJ, 1993.
23. V. Ginzburg, V. Guillemin, and S. Sternberg, *Cobordism Techniques in Symplectic Geometry*, The Carus Mathematical Monographs, Mathematical Association of America, to appear.
24. M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, second edition. Algorithms and Combinatorics, **2**, Springer-Verlag, Berlin, 1993.
25. F. Hirzebruch, *Topological Methods in Algebraic Geometry*, Grundlehren der mathematischen Wissenschaften, **131**, Springer-Verlag, Berlin, 1966.
26. S. Infirri, Lefschetz fixed-point theorem and lattice points in convex polytopes, preprint.

27. R. Kannan, Test sets for integer programs, $\forall\exists$ sentences, in: *Polyhedral Combinatorics (Morristown, NJ, 1989)*, 39–47, DIMACS Ser. Discrete Mathematics and Theoretical Computer Science, 1, American Mathematical Society, Providence, RI, 1990.

28. R. Kannan, Lattice translates of a polytope and the Frobenius problem, *Combinatorica*, **12**(1992), 161–177.

29. R. Kannan, L. Lovász, M. Simonovits, Random walks and an $O^*(n^5)$ volume algorithm for convex bodies, *Random Structures & Algorithms*, **11**(1997), 1–50.

30. J.M. Kantor and A.G. Khovanskii, Integral points in convex polyhedra, combinatorial Riemann-Roch Theorem and generalized Euler-Maclaurin formula, *preprint*, IHES/1992/24, Institut des Hautes Études Scientifiques, Bures-sur-Yvette, 1992.

31. J.M. Kantor and A.G. Khovanskii, Une application du théorème Riemann-Roch combinatoire au polynôme d'Ehrhart des polytopes entiers de $\mathbb{R}^d$, *Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique*, **317**(1993) 501–507.

32. A.G. Khovanskii and A.V. Pukhlikov, Finitely additive measures of virtual polyhedra (Russian), *Algebra i Analiz*, **4**(1992), no. 2, 161–185; translation in *St. Petersburg Mathematical Journal* **4** (1993), no. 2, 337–356.

33. A.G. Khovanskii and A.V. Pukhlikov, The Riemann-Roch theorem for integrals and sums of quasipolynomials on virtual polytopes, (Russian) *Algebra i Analiz* **4** (1992), no. 4, 188–216; *translation in St. Petersburg Mathematical Journal*, **4** (1993), no. 4, 789–812.

34. A.G. Khovanskii and A.V. Pukhlikov, Integral transforms based on Euler characteristic and their applications, *Integral Transforms and Special Functions*, **1**(1993), 19–26.

35. J.C. Lagarias, Point lattices, in: *Handbook of Combinatorics, Vol. 1, 2*, 919–966, Elsevier, Amsterdam, 1995.

36. J. Lawrence, Valuations and polarity, *Discrete & Computational Geometry*, **3**(1988), 307–324.

37. J. Lawrence, Rational-function-valued valuations on polyhedra, in: *Discrete and Computational Geometry (New Brunswick, NJ, 1989/1990)*, 199–208, DIMACS Ser. Discrete Mathematics and Theoretical Computer Science, **6**, American Mathematical Society, Providence, RI, 1991.

38. H.W. Lenstra Jr., Integer programming with a fixed number of variables, *Mathematics of Operations Research*, **8**(1983), 538–548.

39. A.K. Lenstra, H.W. Lenstra Jr., and L. Lovász, Factoring polynomials with rational coefficients, *Mathematische Annalen*, **261** (1982), no. 4, 515–534.

40. L. Lovász, *An Algorithmic Theory of Numbers, Graphs and Convexity*, CBMS-NSF Regional Conference Series in Applied Mathematics, **50**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, 1986.

41. P. McMullen, Non-linear angle-sum relations for polyhedral cones and polytopes, *Mathematical Proceedings of the Cambridge Philosophical Society*, **78**(1975), no. 2, 247–261.

42. P. McMullen, Lattice invariant valuations on rational polytopes, *Archiv der Mathematik* (Basel) , **31** (1978/79), 509–516.

43. P. McMullen, The polytope algebra, *Advances in Mathematics*, **78**(1989), 76–130.

44. P. McMullen, Valuations and dissections, in: *Handbook of Convex Geometry, Vol. A, B*, 933–988, North-Holland, Amsterdam, 1993.
45. P. McMullen and R. Schneider, Valuations on convex bodies, in: *Convexity and its Applications*, 170–247, Birkhäuser, Basel-Boston, Mass., 1983.
46. L.J. Mordell, Lattice points in a tetrahedron and generalized Dedekind sums, *Journal of the Indian Mathematical Society (New Series)*, **15**(1951), 41–46.
47. R. Morelli, A theory of polyhedra, *Advances in Mathematics*, **97**(1993), 1–73.
48. R. Morelli, Translation scissors congruence, *Advances in Mathematics*, **100**(1993), 1–27.
49. R. Morelli, The K-theory of a toric variety, *Advances in Mathematics*, **100**(1993), 154–182.
50. R. Morelli, Pick's Theorem and the Todd class of a toric variety, *Advances in Mathematics*, **100**(1993), 183–231.
51. I. Peeva and B. Sturmfels, Syzygies of codimension 2 lattice ideals, *manuscript*, 1996.
52. G. Pick, *Geometrisches zur Zahlenhre, Naturwissenschaft Zeitschrift Lotos*, Prague, 1899.
53. J.E. Pommersheim, Toric varieties, lattice points and Dedekind sums, *Mathematische Annalen*, **295**(1993), 1–24.
54. J.E. Pommersheim, Products of cycles and the Todd class of a toric variety, *Journal of the American Mathematical Society*, **9**(1996), 813–826.
55. J.E. Pommersheim, Barvinok's algorithm and the Todd class of a toric variety, *Journal of Pure and Applied Algebra*, **117& 118**(1997), 519-533.
56. J.E. Pommersheim and S. Robins, Higher-dimensional Dedekind sums: reciprocity laws, computational complexity, and geometry, *in preparation.*
57. H. Rademacher and E. Grosswald, *Dedekind Sums*, The Carus Mathematical Monographs, No. 16. The Mathematical Association of America, Washington, D.C., 1972.
58. A. Schrijver, *Theory of Linear and Integer Programming*, Wiley-Interscience Series in Discrete Mathematics. A Wiley-Interscience Publication, John Wiley & Sons, Ltd., Chichester, 1986.
59. M.M. Skriganov, Ergodic theory on $SL(n)$, Diophantine approximations and anomalies in the lattice point problem, *Inventiones Mathematicae*, **132**(1998), 1–72.
60. R.P. Stanley, A monotonicity property of h-vectors and h^*-vectors, *European Journal of Combinatorics*, **14** (1993), 251–258.
61. R.P. Stanley, *Enumerative Combinatorics. Vol. 1.*, Cambridge Studies in Advanced Mathematics, **49**, Cambridge University Press, Cambridge, 1997.
62. L.A. Székely and N.C. Wormald, Generating functions for the Frobenius problem with 2 and 3 generators, *Mathematical Chronicle*, **15**(1986), 49–57.
63. D. Zagier, Higher dimensional Dedekind sums, *Mathematische Annalen*, **202**(1973), 149-172.
64. G.M. Ziegler, *Lectures on Polytopes*, Graduate Texts in Mathematics, **152**, Springer-Verlag, New York, 1995.

Alexander Barvinok, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109–1109, USA; e-mail: barvinok@math.lsa.umich.edu

James E. Pommersheim, Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA; email: jamie@math.nmsu.edu