Lattice Points, Polyhedra, and Complexity

Alexander Barvinok

IAS/Park City Mathematics Series
Volume , 2004

Lattice Points, Polyhedra, and Complexity

Alexander Barvinok

Introduction

The central topic of these lectures is efficient counting of integer points in polyhedra. Consequently, various structural results about polyhedra and integer points are ultimately discussed with an eye on computational complexity and algorithms. This approach is one of many possible and it suggests some new analogies and connections. For example, we consider unimodular decompositions of cones as a higher-dimensional generalization of the classical construction of continued fractions. There is a well recognized difference between the theoretical computational complexity of an algorithm and the performance of a computational procedure in practice. Recent computational advances [**L+04**], [**V+04**] demonstrate that many of the theoretical ideas described in these notes indeed work fine in practice. On the other hand, some other theoretically efficient algorithms look completely "unimplementable", a good example is given by some algorithms of [**BW03**]. Moreover, there are problems for which theoretically efficient algorithms are not available at the time. In our view, this indicates the current lack of understanding of some important structural issues in the theory of lattice points and polyhedra. It shows that the theory is very much alive and open for explorations.

Exercises constitute an important part of these notes. They are assembled at the end of each lecture and classified as review problems, supplementary problems, and preview problems.

Review problems ask the reader to complete a proof, to fill some gaps in a proof, or to establish some necessary technical prerequisites. Problems of this kind tend to be relatively straightforward. To be able to complete them is essential for understanding.

Supplementary problems explore various topics in more depth and breadth. Problems of this kind can be harder. They may use some general concepts which

[1]Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109
E-mail address: barvinok@umich.edu

are not formally introduced in the text, but which, nevertheless, are likely to be familiar to the reader.

Preview problems address what is going to appear in the following lectures or after the last lecture. The purpose of these problems is to make the reader prepared, to the extent possible, for further developments.

Acknowledgment. I am grateful to Ezra Miller, Vic Reiner, and Bernd Sturmfels, the organizers of the 2004 Graduate Summer School at Park City, for their invitation to give these lectures and for their support. I am grateful to students and researchers who attended the lectures, asked questions, and otherwise showed their interest in the material. It is my pleasure to thank Greg Blekherman for the excellent job of conducting review sessions where the lecture material was discussed and problems were solved. I am indebted to Greg Blekherman and Kevin Woods for reading the first, pre-event, version of the notes and suggesting corrections and improvements.

This work is partially supported by the NSF grant DMS 0400617.

LECTURE 1
Inspirational Examples. Valuations

The theory we are about to describe is inspired by two simple well-known formulas.

Our first inspiration comes from the formula for the sum of the finite geometric series.

Example 1.

$$\sum_{m=0}^{n} x^m = \frac{1-x^{n+1}}{1-x}.$$

We observe that the long polynomial on the left hand side of the equation sums up to a short rational function on the right hand side.

Geometrically, we do the following: we take the interval $[0, n]$, for every integer point m in the interval we write the monomial x^m, and then take the sum over the integer points in the interval, see Figure 1.

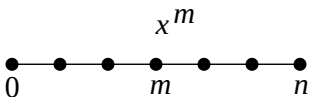

Figure 1. Integer points in the interval

We observe that the thus obtained "long" polynomial (it contains $n+1$ monomials) can be written as a "short" rational function (it is expressed in terms of only 4 monomials).

Naturally, we ask what happens if we replace the interval by something higher-dimensional. Let us, for example, draw a big triangle in the plane, for each integer point $m = (m_1, m_2)$ in the triangle let us write the bivariate monomial $\mathbf{x}^m = x_1^{m_1} x_2^{m_2}$, and then let us try to write the sum over all integer points in the triangle

as some simple rational function in x_1 and x_2, see Figure 2.

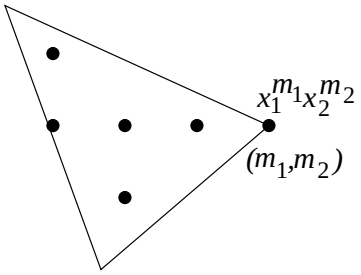

Figure 2. Integer points in the triangle

If the triangle is really large, we get a really long polynomial this way. Later, we will see how to write it as a short rational function.

Our second inspiration comes from the formula for the sum of the infinite geometric series.

Example 2.

$$\sum_{m=0}^{+\infty} x^m = \frac{1}{1-x}.$$

This formula makes sense because the series on the left hand side converges for all $|x| < 1$ to the function on the right hand side. Similarly,

$$\sum_{m=-\infty}^{0} x^m = \frac{1}{1-x^{-1}} = \frac{-x}{1-x}$$

makes sense because the series converges for all $|x| > 1$.

How do we make sense of

$$\sum_{m=-\infty}^{+\infty} x^m \qquad ?$$

This sum does not converge for any x, so we take the easiest route and say that the sum is 0. This may look bizarre but there is some consistence in the way we define the sums: the inclusion-exclusion principle seems to be respected. Indeed, we get the set of all integers if we take all non-negative integers, add all non-positive integers, and subtract 0, as it was double-counted:

$$\sum_{m=-\infty}^{+\infty} x^m = \sum_{m=0}^{+\infty} x^m + \sum_{m=-\infty}^{0} x^m \quad - x^0.$$

This suspiciously agrees with

$$0 = \frac{1}{1-x} + \frac{-x}{1-x} - 1.$$

Geometrically, the real line $\mathbb{R}^1$ is divided into two unbounded rays intersecting in a point. For every region (the two rays, the line, and the point), we construct a rational function so that the sum of x^m over the lattice points in the region converges to that rational function, if converges at all, and the inclusion-exclusion principle is upheld, see Figure 3.

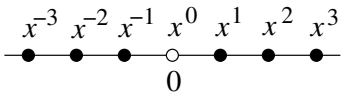

Figure 3. The real line divided into two rays

Naturally, we ask what happens in higher dimensions. Let us draw three lines in general position in the plane: each line splits the plane into two halfplanes, every two lines form four angles, and there are various other regions (one triangle, the whole plane, and some nameless unbounded polygonal regions), see Figure 4.

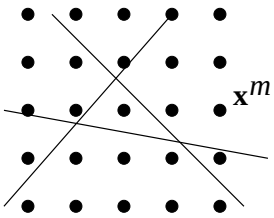

Figure 4. The plane divided into regions

Among those regions, there are regions R where the sum $\sum_{m\in R\cap\mathbb{Z}^2}\mathbf{x}^m$ converges for some $\mathbf{x}$, and there are regions where such a sum would never converge. Can we assign a rational function to every region simultaneously so that each series converges to the corresponding rational function, if converges at all, and the inclusion-exclusion principle is observed?

Later, we will see that such an assignment is indeed possible.

We need some definitions.

Definition 1. The action takes place in Euclidean space $\mathbb{R}^d$, with coordinates $x=(x_1,\dots,x_d)$, the scalar product

$$\langle x,y\rangle=\sum_{i=1}^d x_iy_i \quad \text{for} \quad x=(x_1,\dots,x_d) \quad \text{and} \quad y=(y_1,\dots,y_d),$$

and hence with the integer point lattice $\mathbb{Z}^d\subset\mathbb{R}^d$, consisting of the points x with integer coordinates. A *polyhedron* $P\subset\mathbb{R}^d$ is the set of solutions to finitely many linear inequalities,

$$P=\Big\{x\in\mathbb{R}^d:\quad \sum_{i=1}^d a_{ij}x_j\le b_i,\quad i=1,\dots,m\Big\}.$$

If all a_{ij}, b_i are integers, the polyhedron is *rational*. The main object in these notes is the set $P \cap \mathbb{Z}^d$ of integer points in a rational polyhedron P.

What can we do with polyhedra? The intersection of finitely many (rational) polyhedra is a (rational) polyhedron. The union doesn't have to be but may happen to be a polyhedron. To account for all possible relations among polyhedra, we introduce the *algebra of polyhedra*.

Definition 2. For a set $A \subset \mathbb{R}^d$, let $[A] : \mathbb{R}^d \longrightarrow \mathbb{R}$ be the indicator of A. Thus $[A]$ is the function on $\mathbb{R}^d$ defined by

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

The *algebra of polyhedra* $\mathcal{P}(\mathbb{R}^d)$ is the vector space spanned by the indicators $[P]$ for all polyhedra $P \subset \mathbb{R}^d$. The coefficient field does not matter much: it can be $\mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$.

The *algebra of rational polyhedra* $\mathcal{P}(\mathbb{Q}^d) \subset \mathcal{P}(\mathbb{R}^d)$ is defined similarly as the subspace spanned by the indicators $[P]$ of rational polyhedra P.

Why do we call $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{Q}^d)$ algebras? So far, we defined $\mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}(\mathbb{Q}^d)$ as vector spaces. There is one obvious algebra structure on $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{Q}^d)$. Namely, let $f, g : \mathbb{R}^d \longrightarrow \mathbb{R}$ be functions from the algebras. Then we can define their point-wise product $h = fg$ by $h(x) = f(x)g(x)$. It is immediate to check that h indeed lies in the corresponding algebra. There is a less obvious though more interesting algebra structure on $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{Q}^d)$, see Supplementary Problem 2 in Lecture 2.

Another observation: as long as $d > 0$, the indicators $[P]$ of (rational) polyhedra $P \subset \mathbb{R}^d$ *do not* form a basis of $\mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}(\mathbb{Q}^d)$, because they are linearly dependent. This is what makes the theory interesting.

Valuations

Let V be a vector space. A linear transformation $\mathcal{P}(\mathbb{R}^d), \mathcal{P}(\mathbb{Q}^d) \longrightarrow V$ is called a *valuation*. Basically, this course is about the existence and properties of one particular valuation $\mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathbb{C}(x_1, \ldots, x_d)$, where $\mathbb{C}(x_1, \ldots, x_d)$ is the space of d-variate rational functions. We saw a glimpse of this valuation in Examples 1 and 2.

To warm up, we introduce one of the simplest and most useful valuations.

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

Sketch of proof. Uniqueness of χ, if it exists, is clear. Thus we have to establish existence. We use induction on the dimension d. If $d = 0$, we define $\chi(f) = f(0)$ and it works.

Suppose that $d > 0$. First, we prove the existence of χ on the subspace of $\mathcal{P}(\mathbb{R}^d)$ spanned by the indicators of bounded polyhedra, also known as polytopes. Let us slice $\mathbb{R}^d$ into copies of $\mathbb{R}^{d-1}$ by the value of the last coordinate of a point. That is, we define H_t to be the hyperplane $x_d = t$. Then H_t looks like $\mathbb{R}^{d-1}$ and by the induction hypothesis there is the Euler characteristic χ_t there. Given a function

$f \in \mathcal{P}(\mathbb{R}^d)$, we define its restriction f_t onto H_t. One can easily check that if f is a linear combination of indicators of bounded polyhedra in $\mathbb{R}^d$ then f_t is a linear combination of indicators of bounded polyhedra in H_t. Hence, we can define $\chi_t(f_t)$. Now, the key observation is that the one-sided limit

$$\lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon})$$

always exists and that for all but finitely many t's it is equal to $\chi_t(f_t)$. In fact, if

$$f = \sum_i \alpha_i [P_i],$$

then

$$\lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon}) = \chi_t(f_t)$$

unless t is the minimum value of the last coordinate on one of the polyhedra P_i in the support of f, see Figure 5.

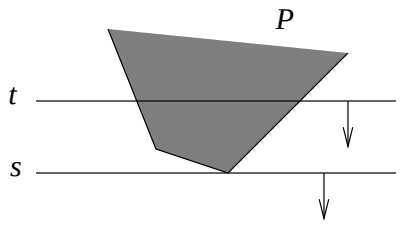

Figure 5. Example: for $f = [P]$, we have $\lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon}) = \chi_t(f_t) = 1$ and $0 = \lim_{\epsilon \longrightarrow +0} \chi_{s-\epsilon}(f_{s-\epsilon}) \neq 1 = \chi_s(f_s)$

This allows us to define

$$\chi(f) = \sum_{t \in \mathbb{R}} \Big(\chi_t(f_t) - \lim_{\epsilon \longrightarrow +0} \chi_{t-\epsilon}(f_{t-\epsilon}) \Big).$$

Although the sum is infinite, only finitely many terms are non-zero.

One can check that χ satisfies the required properties.

Now, we extend χ to the whole algebra $\mathcal{P}(\mathbb{R}^d)$. Let us take P_t to be the cube $|x_i| \leq t$ for $i = 1, \ldots, d$ and let us define

$$\chi(f) = \lim_{t \longrightarrow +\infty} \chi(f \cdot [P_t]) \quad \text{for} \quad f \in \mathcal{P}(\mathbb{R}^d).$$

□

Problems

Review problems.

1. Let $A_1, \ldots, A_n \subset \mathbb{R}^d$ be sets. Prove the inclusion-exclusion formula

$$\left[\bigcup_{i=1}^n A_i\right] = \sum_I (-1)^{|I|-1} \left[\bigcap_{i \in I} A_i\right],$$

where the sum is taken over all non-empty subsets $I \subset \{1, \ldots, n\}$ and $|I|$ is the cardinality of I.

2. Fill in the gaps in the proof of Theorem 1.

3. Show that the Euler characteristic can be extended to the space spanned by the indicators $[A]$ of closed convex sets $A \subset \mathbb{R}^d$ so that $\chi([A]) = 1$ if A is a non-empty closed convex set (a set A is called *convex* if, for every pair of points $x, y \in A$ it contains the interval $[x, y] = \{\alpha x + (1-\alpha)y : \ 0 \leq \alpha \leq 1\}$).

A supplementary problem.

1. Let $P \subset \mathbb{R}^d$ be a bounded polyhedron with a non-empty interior $\operatorname{int} P$. Show that $[\operatorname{int} P] \in \mathcal{P}(\mathbb{R}^d)$ and that $\chi([\operatorname{int} P]) = (-1)^d$. Deduce the Euler-Poincaré formula: if P is a d-dimensional polytope (bounded polyhedron), then $\sum_{k=0}^d (-1)^k f_k = 1$, where f_k is the number of k-dimensional faces of P (including the polytope itself).

Preview problems.

1. Let $P \subset \mathbb{R}^d$ be a polyhedron and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation. Prove that $T(P)$ is a polyhedron.

2. We know that whenever there is an Euler characteristic, there must be an underlying cohomology theory. What is the underlying cohomology theory for the Euler characteristic in Theorem 1?

One problem is that the Euler characteristic of Theorem 1 is not a topological invariant: we have $\chi([A]) = 1 \neq -1 = \chi([B])$, where A is a line and B is an open interval. Hence the underlying cohomology theory must somehow distinguish between bounded and unbounded sets.

Remarks

Theorem 1 and its proof is due to H. Hadwiger, see also Section I.7 of [**Ba02**] for more detail.

LECTURE 2
Identities in the Algebra of Polyhedra

What can we do with polyhedra? One important observation is that the image of a polyhedron under a linear transformation is a polyhedron.

Theorem 1. *Let $P \subset \mathbb{R}^d$ be a polyhedron and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation. Then $T(P) \subset \mathbb{R}^k$ is a polyhedron. Furthermore, if P is a rational polyhedron and T is a rational linear transformation (that is, the matrix of T is rational), then $T(P)$ is a rational polyhedron.*

The crucial step in the proof. Let us consider the following particular case: $k = d-1$ and T is the projection onto the first $(d-1)$ coordinates: $(x_1, \dots, x_d) \longmapsto (x_1, \dots, x_{d-1})$. Suppose that the polyhedron P is defined by a system of linear inequalities:

$$\sum_{j=1}^{d} a_{ij} x_j \le b_i \quad \text{for} \quad i = 1, \dots, m.$$

Let us look at the coefficients of x_d.

Let $I_+ = \{i : \ a_{id} > 0\}$, $I_- = \{i : \ a_{id} < 0\}$, and $I_0 = \{i : \ a_{id} = 0\}$. Then a point $y = (x_1, \dots, x_{d-1})$ belongs to $T(P)$ if and only if

$$\sum_{j=1}^{d-1} a_{ij} x_j \le b_j \quad \text{for} \quad i \in I_0 \tag{1}$$

and there exists x_d such that

$$\begin{aligned} x_d &\le \frac{b_i}{a_{id}} - \sum_{j=1}^{d-1} \frac{a_{ij}}{a_{id}} x_j \quad \text{for} \quad i \in I_+ \\ x_d &\ge \frac{b_i}{a_{id}} - \sum_{j=1}^{d-1} \frac{a_{ij}}{a_{id}} x_j \quad \text{for} \quad i \in I_- \end{aligned} \tag{2}$$

Conditions (1) are some linear inequalities needed to describe $T(P)$, but not all of them. We get the complete set of linear inequalities by majorizing every lower bound by every upper bound in (2), see Figure 6:

$$\frac{b_{i_1}}{a_{i_1 d}} - \sum_{j=1}^{d-1} \frac{a_{i_1 j}}{a_{i_1 d}} x_j \ge \frac{b_{i_2}}{a_{i_2 d}} - \sum_{j=1}^{d-1} \frac{a_{i_2 j}}{a_{i_2 d}} x_j \quad \text{for every pair} \quad i_1 \in I_+, i_2 \in I_-.$$

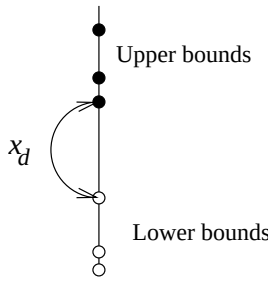

Figure 6. The interval for x_d is obtained by majorizing every lower bound by every upper bound

Thus we perform a step of the procedure known as the *Fourier-Motzkin elimination.* □

Linear transformations preserve linear relations among indicators of polyhedra.

Theorem 2. *Let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation. Then there exists a linear transformation $\mathcal{T} : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^k)$ such that $\mathcal{T}[P] = [T(P)]$ for every polyhedron $P \subset \mathbb{R}^d$.*

Proof. Let us define the "kernel" $G : \mathbb{R}^d \times \mathbb{R}^k \longrightarrow \mathbb{R}$ by

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

Let us choose $f \in \mathcal{P}(\mathbb{R}^d)$. We must define $h \in \mathcal{P}(\mathbb{R}^k)$ such that $\mathcal{T}(f) = h$. To this end, for every $y \in \mathbb{R}^k$, we define a function $g_y \in \mathcal{P}(\mathbb{R}^d)$ by $g_y(x) = G(x,y)f(x)$. One can check that $g_y \in \mathcal{P}(\mathbb{R}^d)$. Hence we can apply the Euler characteristic χ to g_y. We let $h(y) = \chi(g_y)$. Thus we got a function $h : \mathbb{R}^k \longrightarrow \mathbb{R}$. Next, one should check that if $f = [P]$ then $h = [T(P)]$. It follows that if $f \in \mathcal{P}(\mathbb{R}^d)$ then $h \in \mathcal{P}(\mathbb{R}^k)$. We conclude that $\mathcal{T}(f) = h$ defines the required linear transformation. □

We call $G(x,y)$ the "kernel" to underline a certain similarity between our construction and the standard construction of various integral operators between functional spaces in analysis. In analysis, we often construct a linear transformation which transforms a function $f : X \longrightarrow \mathbb{R}$ into a function $h : Y \longrightarrow \mathbb{R}$ by choosing an appropriate kernel $K(x,y) : X \times Y \longrightarrow \mathbb{R}$ and defining $h(y) = \int K(x,y)f(x)\, d\mu(x)$ for some measure μ on X. In polyhedral combinatorics, we can construct some interesting linear operators $\mathcal{A} : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^k)$ by choosing an appropriate "kernel" $K(x,y) : \mathbb{R}^d \times \mathbb{R}^k \longrightarrow \mathbb{R}$ and letting $\mathcal{A}(f) = h$, where $h(y) = \chi(g_y)$ for $g_y(x) = K(x,y)f(x)$. The similarity is partially explained by the observation that one can think of the Euler characteristic as a finitely-additive measure on $\mathbb{R}^d$ that is a "combinatorialization" of the Lebesgue measure. In analysis, we want to know how large is a given set and the Lebesgue measure tells us that. In polyhedral combinatorics, we just want to know whether a given polyhedron is non-empty, and the Euler characteristic tells that.

It follows from Theorem 2 that whenever we have a linear relation $\sum_{i=1}^m \alpha_i [P_i] = 0$ among the indicator functions of polyhedra, the same relation $\sum_{i=1}^m \alpha_i [T(P_i)] = 0$ holds for their images under a linear transformation. This is obvious for invertible transformations T but strarting to look less obvious for projections, see Figure 7 for a simple example.

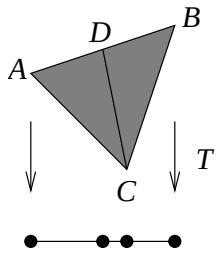

Figure 7. We have $[ABC] = [ACD] + [CBD] - [CD]$ and $[T(ABC)] = [T(ACD)] + [T(CBD)] - [T(CD)]$

Now we need to take a closer look at polyhedra. Some polyhedra have *vertices*, some don't.

Definition 1. Let $P \subset \mathbb{R}^d$ be a polyhedron. A point $v \in P$ is called a *vertex* of P if whenever $v = (x+y)/2$ for some $x, y \in P$, we must have $x = y = v$. If v is a point in P, we define the *tangent cone* of P at v as follows:

$$\mathrm{co}(P, v) = \Big\{ x \in \mathbb{R}^d : \quad \epsilon x + (1-\epsilon) v \in P \quad \text{for all sufficiently small} \quad \epsilon > 0 \Big\}.$$

Figure 8 shows what tangent cones may look like.

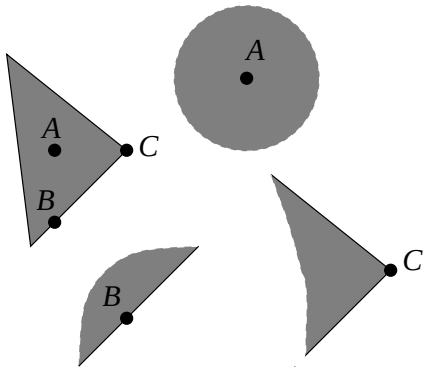

Figure 8. A polyhedron and its tangent cones

Not all polyhedra have vertices. In fact, a non-empty polyhedron has a vertex if and only if it does not contain a line.

Definition 2. We say that a polyhedron P contains a *line* if there are points x and y such that $y \neq 0$ and $x + ty \in P$ for all $t \in \mathbb{R}$. Finally, let $\mathcal{P}_0(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}_0(\mathbb{Q}^d) \subset \mathcal{P}(\mathbb{Q}^d)$ be the subspace spanned by the indicators of (rational) polyhedra that contain lines.

It turns out that modulo polyhedra with lines, every polyhedron is just the sum of its tangent cones.

Theorem 3. *Let $P \subset \mathbb{R}^d$ be a polyhedron. Then there is a $g \in \mathcal{P}_0(\mathbb{R}^d)$ such that*

$$[P] = g + \sum_v [\mathrm{co}(P, v)],$$

where the sum is taken over all vertices v of P. If P is a rational polytope then we can choose $g \in \mathcal{P}_0(\mathbb{Q}^d)$.

A plausible argument. We don't really prove this important theorem, although we come very close. We start by showing that the theorem is not obviously false.

We notice that if P is non-empty and does not contain vertices then P contains a line and hence we can choose $g = [P]$.

Suppose we have been sloppy and included in the sum not only all vertices v of P but also some non-vertices $v \in P$. No harm done: if $v \in P$ is a non-vertex then $\mathrm{co}(P, v)$ contains a line and so we just have to adjust g. This shows that the formula is robust enough.

Suppose that the theorem holds for some polyhedron $P \subset \mathbb{R}^d$ and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a sufficiently generic linear transformation. We claim that the theorem holds for the image $T(P)$. Indeed, by Theorem 2 the transformation T gives rise to the transformation $\mathcal{T}$ on the algebra of polyhedra. Let us apply $\mathcal{T}$ to both sides of the identity. We have $\mathcal{T}[P] = [T(P)]$ and $\mathcal{T}[\mathrm{co}(P, v)] = [T(\mathrm{co}(P, v))] = [\mathrm{co}(T(P), T(v))]$. We have to be somewhat careful with g: we know that g is a linear combination of indicators of polyhedra with lines. If we are unlucky, the kernel of T may "eat up" some of those lines and $\mathcal{T}(g)$ will not lie in $\mathcal{P}_0(\mathbb{R}^k)$. This is the reason why we chose T to be "generic". Thus if we prove the theorem for some "model" polyhedra P, we can extend it (with some care) to polyhedra obtained from P by linear transformations.

Now, we show that the result holds for a *simplex*, which we define as a compact polyhedron $\Delta \subset \mathbb{R}^d$ that is the intersection of $d + 1$ sufficiently generic halfspaces $H_1, \ldots, H_{d+1}$. We notice that $[H_1 \cup \ldots \cup H_{d+1}] = [\mathbb{R}^d]$ and expanding $[H_1 \cup \ldots \cup H_{d+1}]$ by the inclusion-exclusion formula we represent $[\mathbb{R}^d]$ as the alternating sum of the indicators $[H_{i_1} \cap \ldots \cap H_{i_k}]$ of intersections of halfspaces. All such intersections contain lines except for the simplex $\Delta = [H_1 \cap \ldots \cap H_{d+1}]$ itself (the intersection of all $d + 1$ halfspaces) and the tangent cones $[H_1 \cap \ldots \cap H_{i-1} \cap H_{i+1} \cap \ldots \cap H_{d+1}]$

(the intersections of all but one halfspace) at the vertices of Δ, see Figure 9.

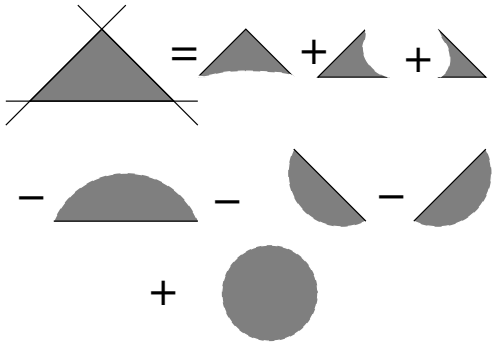

Figure 9. A triangle is the sum of the angles at its vertices minus the halfplanes based on its sides plus the whole plane

It follows now that the result holds for all projections of simplices, that is for polytopes (bounded polyhedra). To obtain the formula for a general polyhedron, one needs some structural results about unbounded polyhedra, namely that every unbounded polyhedron is the Minkowski sum of its *recession cone* and a polytope, see Review Problem 11 and Supplementary Problem 3. □

Definition 3. Let $A \subset \mathbb{R}^d$ be a non-empty set. The set

$$A^\circ = \Big\{ y \in \mathbb{R}^d : \ \langle x, y \rangle \leq 1 \quad \text{for all} \quad x \in A \Big\}$$

is called the *polar* of A.

It is easy to see that A° is a non-empty closed convex set containing the origin. The *Bipolar Theorem* asserts that $(A^\circ)^\circ = A$ provided A is a closed convex set containing the origin. One can show that if P is a (rational) polyhedron then P° is a (rational) polyhedron, see Figure 10.

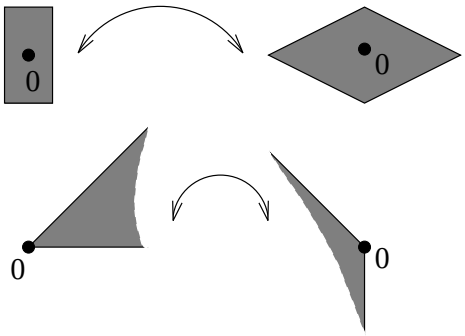

Figure 10. Some (bounded and unbounded) polyhedra and their polars

It is somewhat surprising that the polarity correspondence preserves linear relations among the indicator functions of polyhedra.

Theorem 4. *There exists linear transformations $\mathcal{D} : \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^d)$, $\mathcal{D} : \mathcal{P}(\mathbb{Q}^d) \longrightarrow \mathcal{P}(\mathbb{Q}^d)$, such that $\mathcal{D}[P] = [P^\circ]$ for every non-empty (rational) polyhedron P.*

The idea of the proof. We define $\mathcal{D}$ as a limit of certain operators $\mathcal{D}_\epsilon$. For $\epsilon > 0$, let us define the kernel $G_\epsilon : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ by

$$G_\epsilon(x,y) = \begin{cases} 1 & \text{if } \langle x,y\rangle < 1+\epsilon \\ 0 & \text{otherwise.} \end{cases}$$

For $f \in \mathcal{P}(\mathbb{R}^d)$, $\mathcal{P}(\mathbb{Q}^d)$ and $y \in \mathbb{R}^d$, let $g_{y,\epsilon}(x) = f(x)G_\epsilon(x,y)$. One can check that $g_{y,\epsilon} \in \mathcal{P}(\mathbb{R}^d), \mathcal{P}(\mathbb{Q}^d)$, so we can apply the Euler characteristic χ to $g_{y,\epsilon}$. Let us define $h_\epsilon = \mathcal{D}_\epsilon(f)$ by $h_\epsilon(y) = \chi(g_{y,\epsilon})$. Finally, we define $h = \mathcal{D}(f)$ by $h(y) = \lim_{\epsilon \longrightarrow 0+} h_\epsilon(y)$. One can check then that $\mathcal{D}$ satisfies the desired properties. □

It follows from Theorem 4 that whenever we have a linear identity $\sum_{i=1}^m \alpha_i [P_i] = 0$ among the indicator functions of polyhedra, we have the same identity $\sum_{i=1}^m \alpha_i [P_i^\circ] = 0$ for the indicator functions of their polars, see Figure 11.

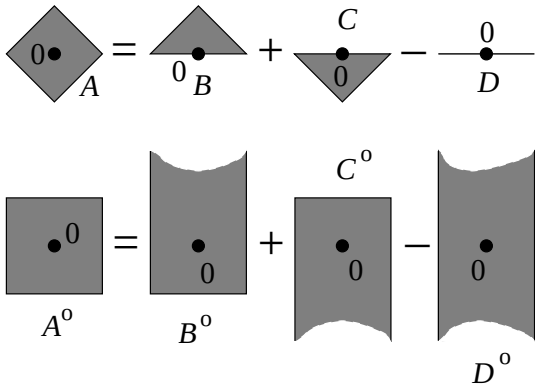

Figure 11. We have $[A] = [B] + [C] - [D]$ and $[A^\circ] = [B^\circ] + [C^\circ] - [D^\circ]$

An important feature of the polarity transform is that P° contains a line if and only if P lies in an affine hyperplane, that is, not full-dimensional. Continuing our analogy with analysis, we can say that in the Euler characteristic based polyhedral combinatorics, the polarity transform $\mathcal{D}$ plays the role akin to that of the Fourier transform in the Lebesgue measure based analysis. An observation in support of this statement can be found in Preview Problem 3.

Problems

Review problems.

1. Complete the proof of Theorem 1.

2. In Theorem 1, suppose that $P \subset \mathbb{R}^d$ is defined by m linear inequalities. Estimate the number of inequalities needed to define $T(P)$.

3. Check the proof of Theorem 2.

4. Let $P \subset \mathbb{R}^d$ be a polyhedron defined by m linear inequalities

$$\sum_{j=1}^{d} a_{ij} x_j \leq b_i \quad \text{for} \quad i = 1, \ldots, m.$$

Let $x \in P$ be a point. We say that the inequality is *active* on x if equality holds at x. Let $a_i = (a_{i1}, \ldots, a_{id})$ be the vector of coefficients of the i-th inequality. Prove that $v \in P$ is a vertex of P if and only if there are at least d inequalities active on v such that their vectors form a basis of $\mathbb{R}^d$.

5. Prove that a polyhedron has finitely many vertices, if any.

6. Let P be a rational polyhedron and let $v \in P$ be a vertex. Prove that v has rational coordinates.

7. Let P be a polyhedron and let $v \in P$ be a point. Prove that $\text{co}(P, v)$ is the polyhedron defined by the inequalities of P that are active on v.

8. Prove that a non-empty polyhedron has a vertex if and only if it does not contain lines.

9. Prove that v is a vertex of P if and only if $\text{co}(P, v)$ does not contain lines.

10. Let $P \subset \mathbb{R}^d$ is a polyhedron, let $v \in P$ be a point, let $T : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be a linear transformation, let $Q = T(P)$, and let $u = T(v)$. Prove that $\text{co}(Q, u) = T(\text{co}(P, v))$.

11. Let $P \subset \mathbb{R}^d$ be a non-empty (rational) polyhedron. Let us define the *recession cone* K_P by

$$K_P = \Big\{ x \in \mathbb{R}^d : \quad y + tx \in P \quad \text{for all} \quad y \in P \quad \text{and all} \quad t \geq 0 \Big\}.$$

Show that K_P is a (rational) polyhedron.

12. Prove that a non-empty polyhedron $P \subset \mathbb{R}^d$ lies in an affine hyperplane if and only if P° contains a line.

Supplementary problems.

1. For sets $A, B \subset \mathbb{R}^d$, we define their *Minkowski sum* $A + B = \big\{ x + y : \ x \in A, y \in B \big\}$. Prove that the Minkowski sum of polyhedra is a polyhedron and that the Minkowski sum of rational polyhedra is a rational polyhedron.

2. Prove that there exists a bilinear operation, called convolution, $\star : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \longrightarrow \mathcal{P}(\mathbb{R}^d)$ such that $[P] \star [Q] = [P + Q]$ for any two polyhedra $P, Q \subset \mathbb{R}^d$. This gives $\mathcal{P}(\mathbb{R}^d)$ another (more interesting) commutative algebra structure. Note that $[0]$ plays the role of the identity, so $f \star [0] = [0] \star f = f$ for all $f \in \mathcal{P}(\mathbb{R}^d)$.

3. Let $P \subset \mathbb{R}^d$ be a non-empty polyhedron not containing lines and let Q be the convex hull of the set of vertices of P. Prove that P can be represented as the Minkowski sum $P = Q + K_P$, where K_P is the recession cone of P, cf. Review Problem 11.

4. Using Problem 3 above, complete the proof of Theorem 3.

5. Suppose that $P \subset \mathbb{R}^d$ is a bounded polyhedron. Prove that $[P]$ is invertible with respect to the convolution operation $\star$ of Problem 2 above : there exists an

$f \in \mathcal{P}(\mathbb{R}^d)$ such that $f \star [P] = [0]$. More precisely, if P is a bounded polyhedron with a non-empty interior $\operatorname{int} P$, we can choose $f = (-1)^d[-\operatorname{int} P]$ (that is, we take the interior of P, reflect it about the origin, and take the indicator of the set we got with the appropriate sign).

6. Let $P \subset \mathbb{R}^d$ be a polyhedron. We say that two points $x, y \in P$ are *equivalent*, if $\operatorname{co}(P, x) = \operatorname{co}(P, y)$. An equivalence class of points in P is just an open face $F \subset P$. For an $x \in F$, we denote $\operatorname{co}(P, x)$ by $\operatorname{co}(P, F)$. Prove the following *Gram-Brianchon Theorem*

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

where the sum is taken over all non-empty faces F of P, including $F = P$.

7. Let $P \subset \mathbb{R}^d$ be a bounded polyhedron (polytope) containing the origin in its interior. For a face F of P, let $P_F = \operatorname{conv}(F, 0)$ be the convex hull of the face F and the origin. Prove that

$$(-1)^{d-1}[P] = \sum_F (-1)^{\dim F} [P_F],$$

where the sum is taken over all faces $F \neq P$ of P, including the empty face (cf. Supplementary Problem 1 to Lecture 1).

8. Prove that the polar of a (rational) polyhedron is a (rational polyhedron) and that $(A^\circ)^\circ = A$ if A is closed, convex, and contains 0.

9. Complete the proof of Theorem 4.

10. Show that if we apply the polarity transform $\mathcal{D}$ to both sides of the identity in Problem 7 above, we get the Gram-Brianchon identity of Problem 6.

Preview problems.

1. A polyhedron $K \subset \mathbb{R}^d$ is called a (polyhedral) *cone* if $0 \in K$ and $\lambda x \in K$ for all $x \in K$ and all $\lambda \geq 0$ (note that the tangent cone of Definition 1 is not necessarily a cone in the sense of this definition, since the vertex of the tangent cone is not necessarily the origin). Prove that if K is a cone then K° is a cone and that $(K^\circ)^\circ = K$.

2. Let $K_1, K_2 \subset \mathbb{R}^d$ be polyhedral cones. Prove that $[K_1 \cap K_2]^\circ = [K_1 + K_2]$, where "+" is the Minkowski sum, see Supplementary Problem 1.

3. Let $\mathcal{D}$ be the transform of Theorem 4 and let $f_1, f_2 \in \mathcal{P}(\mathbb{R}^d)$ be linear combinations of indicator functions of polyhedral cones. Prove that $\mathcal{D}(f_1 f_2) = \mathcal{D}(f_1) \star \mathcal{D}(f_2)$, where $\star$ is the convolution operation from Supplementary Problem 2.

Remarks

For the Fourier-Motzkin elimination (Theorem 1), see Sections I.9 of [**Ba02**] and Sections 1.2-1.3 of [**Zi95**]. A nice exposition of the Euler characteristic and the theory of valuations is given in [**KR97**]. Much of the material of this lecture can be found in [**Ba02**]: Section II.4-5 (vertices of polyhedra), Section IV.1 (polarity), Section VIII.4 (tangent cones). Analogies between integral operators in the classical analysis and valuations are drawn, for example, in [**KP93**].

LECTURE 3
Generating Functions and Cones. Continued Fractions

Now we turn to integer points. For an integer point $m = (m_1, \dots, m_d)$, we introduce the monomial $\mathbf{x}^m = x_1^{m_1} \cdots x_d^{m_d}$ in d complex variables $\mathbf{x} = (x_1, \dots, x_d)$. Given a set $S \subset \mathbb{R}^d$, we consider the sum

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

Our goal is to find a reasonably short expression for this sum as a rational function in $\mathbf{x}$. Our inspiration is the formula

$$\sum_{m=0}^{+\infty} x^m = \frac{1}{1-x} \quad \text{for} \quad |x| < 1.$$

Here is an obvious multivariate generalization of the formula.

Example 1. Let $\mathbb{R}_+^d$ be the non-negative orthant, that is the set of points with all coordinates non-negative. We have

$$\sum_{m \in \mathbb{R}_+^d \cap \mathbb{Z}^d} \mathbf{x}^m = \left(\sum_{m_1=0}^{+\infty} x_1^{m_1} \right) \cdots \left(\sum_{m_d=0}^{+\infty} x_d^{m_d} \right)$$

$$= \prod_{i=1}^{d} \frac{1}{1-x_i} \quad \text{provided} \quad |x_i| < 1 \quad \text{for} \quad i = 1, \dots, d.$$

In general, we say that $f(S, \mathbf{x})$ is defined by a particular rational function if there is a non-empty open set $U \subset \mathbb{C}^d$ such that for all $\mathbf{x} \in U$ the defining series for $f(S, \mathbf{x})$ converges absolutely to that rational function and the convergence is uniform on compact subsets of U. In all cases we encounter, only existence of such an U, but not its precise shape will be of importance.

Our next step is less obvious. What if the orthant gets somewhat "skewed"?

Definition 1. Let $u_1, \dots, u_d \in \mathbb{Z}^d$ be linearly independent integer vectors. The *simple rational cone* generated by $u_1, \dots, u_d$ is the set

$$K = K(u_1, \dots, u_d) = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad \alpha_i \geq 0 \quad \text{for} \quad i = 1, \dots, d \Big\}.$$

The *fundamental parallelepiped* of $u_1, \dots, u_d$ is the set

$$\Pi = \Pi(u_1, \dots, u_d) = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad 0 \le \alpha_i < 1 \quad \text{for} \quad i = 1, \dots, d \Big\}.$$

Note that the parallelepiped is "semi-open", see Figure 12.

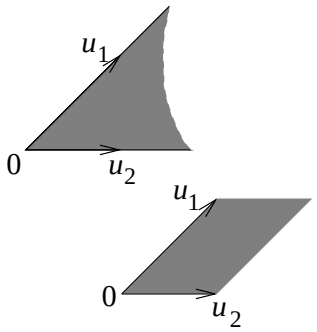

Figure 12. A simple rational cone and its fundamental parallelepiped

Clearly, if we scale vectors $u_1, \dots, u_d$: $u_i := k_i u_i$ for some positive integers k_i, the cone K will not change (although the parallelepiped Π will). This is all the freedom we have in choosing $u_1, \dots, u_d$ for a given K.

Let $u_i^*, i = 1, \dots, d$ be the vectors defined by $\langle u_i, u_j^* \rangle = -\delta_{ij}$. Then

$$K = \Big\{ x : \quad \langle x, u_i^* \rangle \le 0 \quad \text{for} \quad i = 1, \dots, d \Big\},$$

from which it follows that simple rational cones are rational polyhedra.

Theorem 1. *For a simple rational cone $K = K(u_1, \dots, u_d)$, we have*

$$f(K, \mathbf{x}) = \left(\sum_{m \in \Pi \cap \mathbb{Z}^d} \mathbf{x}^m \right) \prod_{i=1}^{d} \frac{1}{1 - \mathbf{x}^{u_i}}.$$

Proof. The proof consists of the observation that *every* point $m \in K \cap \mathbb{Z}^d$ can be *uniquely* written as $m = m_1 + m_2$, where $m_1 \in \Pi \cap \mathbb{Z}^d$ and m_2 is a non-negative integer combination of $u_1, \dots, u_d$. Indeed, since u lies in the cone K, it can be written in the form

$$m = \sum_{i=1}^{d} \alpha_i u_i \quad \text{for some real numbers} \quad \alpha_i \ge 0.$$

Let $\lfloor \alpha \rfloor$ denote the largest integer not exceeding α (a.k.a the *integer part* of α) and let $\{\alpha\} = \alpha - \lfloor \alpha \rfloor$ (the *fractional part* of α). Then

$$m_1 = \sum_{i=1}^{d} \{\alpha_i\} u_i \quad \text{and} \quad m_2 = \sum_{i=1}^{d} \lfloor \alpha_i \rfloor u_i.$$

To prove uniqueness, suppose that we have two decompositions $m = m_1 + m_2$ and $m = m_1' + m_2'$, where m_1 and m_2 are integer points from the parallelepiped Π and m_2 and m_2' are non-negative integer combinations of $u_1, \dots, u_d$. Then we can write $m_1 - m_1' = m_2' - m_2$, from which $m_1 - m_1'$ is an integer combination of $u_1, \dots, u_d$. However, since $m_1, m_1' \in \Pi$, we should be able to write

$$m_1 - m_1' = \sum_{i=1}^{d} \beta_i u_i \quad \text{where} \quad -1 < \beta_i < 1 \quad \text{for} \quad i = 1, \dots, d.$$

Thus we must have $\beta_i = 0$ and $m_1 = m_1'$, $m_2 = m_2'$.

It remains to show that there is some non-empty open set $U \subset \mathbb{C}^d$ of $\mathbf{x}$ for which the series

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

converges absolutely and uniformly on compact subsets of U. Since $u_1, \dots, u_d$ are linearly independent, we can find a vector $c = (c_1, \dots, c_d)$, such that $\langle c, u_i \rangle < 0$ for $i = 1, \dots, d$, where $\langle x, y \rangle = x_1 y_1 + \ldots + x_d y_d$ is the standard scalar product in $\mathbb{R}^d$. Let $\mathbf{x}_0 = (e^{c_1}, \dots, e^{c_d})$. Then for all $\mathbf{x}$ in a sufficiently small neighborhood U of $\mathbf{x}_0$, the series converges as desired. Since the product $\prod_{i=1}^{d} (1 - \mathbf{x}^{u_i})^{-1}$ encodes the sum of $\mathbf{x}^m$ over all m that are non-negative integer combinations of $u_1, \dots, u_d$ (cf. Example 1), the proof follows. □

Theorem 1 provides us with a finite formula for an infinite series, but there is still something unsatisfactory about it. Namely, the sum over integer points in the fundamental parallelepiped is not very explicit and, although finite, can be quite large. However, although we can't predict which integer points lie in the parallelepiped, we can tell the number of such points exactly.

Theorem 2. *The number of integer points in the fundamental parallelepiped is equal to the volume of the parallelepiped.*

Sketch of proof. Let Λ be the set of all integer combinations of $u_1, \dots, u_d$:

$$\Lambda = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad \alpha_i \in \mathbb{Z} \quad \text{for} \quad i = 1, \dots, d \Big\}.$$

Let us consider all translates $\Pi + u$ with $u \in \Lambda$. We claim that the translations $\Pi + u : \ u \in \Lambda$ cover $\mathbb{R}^d$ without overlapping. The proof can be extracted from the proof of Theorem 1. Let us take a sufficiently large, regular looking region $X \subset \mathbb{R}^d$ (say, a ball of a large radius), and let us count integer points in X. On one hand, we can approximate the number of integer points in X by the volume $\operatorname{vol} X$ of X. One the other hand, the set is covered by roughly $(\operatorname{vol} X)/(\operatorname{vol} \Pi)$ translations of the parallelepiped Π and each translation contains the same number of integer points. Hence we must have $|\Pi \cap \mathbb{Z}^d| = \operatorname{vol} \Pi$. □

For a simple illustration of Theorem 2, see Figure 13.

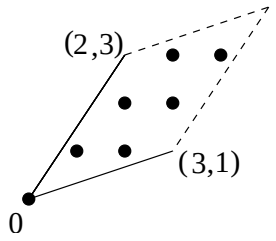

Figure 13. The number of integer points in a fundamental parallelogram is equal to the area of the parallelogram

This leads us to the following crucial definition.

Definition 2. Let $u_1, \ldots, u_d \in \mathbb{Z}^d$ be linearly independent vectors and let K be the simple cone generated by $u_1, \ldots, u_d$. We say that K is *unimodular* if the volume of the fundamental parallelepiped Π is 1. Equivalently, K is unimodular if the origin is the unique integer point in Π. Equivalently,

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

One of our goals is to devise an efficient procedure of *decomposing* a given simple cone into a certain *combination* of unimodular cones. The first non-trivial case is $d = 2$ (every 1-dimensional cone is unimodular) and there such a procedure has long been known.

Continued fractions

Let us choose a number $a \in \mathbb{R}$. The following procedure produces what is called the *continued fraction* expansion $[a_0; a_1, \ldots, a_n, \ldots]$ of a. First, we write

$$a = \lfloor a \rfloor + \{a\} \quad \text{and let} \quad a_0 = \lfloor a \rfloor.$$

Now, if $\{a\} = 0$, we stop. Otherwise, $0 < \{a\} < 1$, we let $b = 1/\{a\}$, so $b > 1$. We write

$$b = \lfloor b \rfloor + \{b\} \quad \text{and let} \quad a_1 = \lfloor b \rfloor.$$

If $\{b\} = 0$ we stop. Otherwise, we let new $b := 1/\{\text{old } b\}$, and continue. In the end, we get the expansion

$$a = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ldots}}}$$

The expansion can be finite (if a is rational) or infinite (if a is irrational). We define the k-th *convergent* $[a_0; a_1, \dots, a_k]$ by cutting the expansion at a_k. For example, the 4-th convergent $[a_0; a_2, a_3, a_4]$ is

$$a = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4.}}}}$$

As an example, let us compute the continued fractions expansion of $a = 164/31$:

$$\frac{164}{31} = 5 + \frac{9}{31} = 5 + \cfrac{1}{3 + \cfrac{4}{9}} = 5 + \cfrac{1}{3 + \cfrac{1}{2 + \cfrac{1}{4.}}}$$

Hence $164/31 = [5; 3, 2, 4]$. Now we compute the convergents:

$$[5; 3, 2] = 5 + \cfrac{1}{3 + \cfrac{1}{2}} = \frac{37}{7}, \quad [5; 3] = 5 + \frac{1}{3} = \frac{16}{3}, \quad [5] = \frac{5}{1}.$$

Computing $f(K, \mathbf{x})$ for 2-dimensional cones

Continued fractions can be applied to obtain short formulas for $f(K, \mathbf{x})$, where $K \subset \mathbb{R}^2$ is a simple cone. Instead of developing a comprehensive theory, we give one computational example.

Let $K \subset \mathbb{R}^2$ be the cone generated by the vectors $(1, 0)$ and $(31, 164)$. The volume of the fundamental parallelepiped is 164, so the formula for $f(K, \mathbf{x})$ provided by Theorem 1 would contain a sum of 164 monomials. We will find a shorter formula, and, moreover, will compute it by hand.

First, we compute the continued fraction expansion of $164/31 = [5; 3, 2, 4]$ and the convergents $[5] = 5/1$, $[5; 3] = 16/3$, $[5; 3, 2] = 37/7$, cf. the above example.

Now, we do some "surgery" on cones. Let us consider the following cones, given by their generators

K_0	generated by	$(1, 0)$	and	$(0, 1)$;	
K_1	generated by	$(1, 0)$	and	$(1, 5)$;	
K_2	generated by	$(1, 0)$	and	$(3, 16)$;	
K_3	generated by	$(1, 0)$	and	$(7, 37)$;	and, finally,
K_4	generated by	$(1, 0)$	and	$(31, 164)$.	

We observe that K_0 is a unimodular cone with

$$f(K_0, \mathbf{x}) = \frac{1}{(1 - x_1)(1 - x_2)},$$

while we are trying to compute $f(K_4, \mathbf{x})$.

The crucial observation is that to pass from K_i to K_{i+1} we have either to "cut" from K_i a unimodular cone (if i is even) or to "paste" to K_i a unimodular cone (i is odd), see Figure 14.

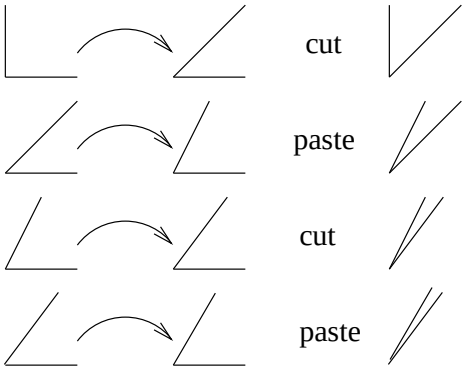

Figure 14. Cutting and pasting unimodular cones

Hence, starting with K_0, we

cut the unimodular cone generated by	$(0,1)$	and	$(1,5)$;
paste the unimodular cone generated by	$(1,5)$	and	$(3,16)$;
cut the unimodular cone generated by	$(3,16)$	and	$(7,37)$;
paste the unimodular cone generated by	$(7,37)$	and	$(31,164)$

to finally get K_4. Taking into account "boundary effects" (when we cut and paste, some points on the boundary get double-counted), which, luckily, cancel each other, we get:

$$\begin{aligned} f(K,\mathbf{x}) =& \frac{1}{(1-x_1)(1-x_2)} - \frac{1}{(1-x_2)(1-x_1x_2^5)} + \frac{1}{1-x_1x_2^5} \\ &+ \frac{1}{(1-x_1x_2^5)(1-x_1^3x_2^{16})} - \frac{1}{1-x_1x_2^5} \\ &- \frac{1}{(1-x_1^3x_2^{16})(1-x_1^7x_2^{37})} + \frac{1}{1-x_1^7x_2^{37}} \\ &+ \frac{1}{(1-x_1^7x_2^{37})(1-x_1^{31}x_2^{164})} - \frac{1}{1-x_1^7x_2^{37}}, \end{aligned}$$

so finally,

$$\begin{aligned} f(K,\mathbf{x}) =& \frac{1}{(1-x_1)(1-x_2)} - \frac{1}{(1-x_2)(1-x_1x_2^5)} + \frac{1}{(1-x_1x_2^5)(1-x_1^3x_2^{16})} \\ &- \frac{1}{(1-x_1^3x_2^{16})(1-x_1^7x_2^{37})} + \frac{1}{(1-x_1^7x_2^{37})(1-x_1^{31}x_2^{164})}. \end{aligned}$$

The formula is reasonably short.

Given an arbitrary 2-dimensional rational cone generated by $u_1, u_2 \in \mathbb{Z}^2$, we can always change the coordinates by applying a linear transformation which preserves $\mathbb{Z}^2$ so that u_2 becomes equal to $(1,0)$. Suppose that $u_1 = (q,p)$ for integers p and $q > 0$. To compute the generating function $f(K,\mathbf{x})$, we compute the continued fraction expansion of p/q and obtain K by cutting and pasting the unimodular cones computed from the convergents of p/q. If the k-th convergent is p_k/q_k, we cut or paste, depending on the parity of $k > 1$, the cone generated by (q_k, p_k) and (q_{k-1}, p_{k-1}), which is always unimodular, see Review Problems 4 and 5.

The computational complexity

Let $K \subset \mathbb{R}^2$ be the cone generated by $u_1 = (1,0)$ and $u_2 = (q,p)$ as above. The fundamental parallelepiped of K contains $|p|$ points, so if we compute $f(K,\mathbf{x})$ by the formula of Theorem 1, the resulting rational function will contain $|p|$ terms. If, instead, we use the continued fractions method, we get an expression for $f(K,\mathbf{x})$ containing about $\log(\min(|p|,|q|) + O(1)$ terms. For large $|p|$, the difference is quite significant. Looking more closely, we observe that to *define* the cone K, that is, to write the coordinates of its generators, we need about $\log|p| + \log|q| + O(1)$ bits (or digits) since to write an integer a we need about $\log|a| + O(1)$ bits (or digits). Thus we say that the *input size* of the problem of computing $f(K,\mathbf{x})$ is about $\log|p| + \log|q| + O(1)$. The number of operations required to compute $f(K,\mathbf{x})$ via continued fractions is about $O(\log^2|p| + \log^2|q| + 1)$, that is, bounded by a *polynomial* in the input size. In contrast, the number operations required to compute $f(K,\mathbf{x})$ via Theorem 1 (and even to write down the answer) is *exponential* in the input size of K. In Lecture 5, for any dimension d (fixed in advance), we present a polynomial time algorithm, which, given a rational cone $K \subset \mathbb{R}^d$ as an input, computes $f(K,\mathbf{x})$ as a rational function.

Problems

Review problems.

1. Check the proof of Theorem 1.

2. Make the proof of Theorem 2 rigorous.

3. Let K be the 2-dimensional simple cone generated by $u_1 = (1,0)$ and $u_2 = (1,n)$ for some positive integer n. Compute $f(K,\mathbf{x})$.

4. Let $[a_0; a_1, \ldots, a_n \ldots]$ be the continued fraction expansion of a real number a and let $p_k/q_k = [a_0; a_1, \ldots, a_k]$ be the k-th convergent (we assume that p_k and q_k are comprime). Prove that for $k \geq 2$

$$p_k = a_k p_{k-1} + p_{k-2} \quad \text{and} \quad q_k = a_k q_{k-1} + q_{k-2}.$$

Deduce that

$$p_{k-1}q_k - p_k q_{k-1} = (-1)^{k-1} \quad \text{for} \quad k \geq 0.$$

5. Justify the procedure of computing $f(K,\mathbf{x})$ for the cone K generated by $(1,0)$ and (q,p) via continued fractions.

6. Let $K \subset \mathbb{R}^d$ be the set defined by

$$K = \Big\{ x \in \mathbb{R}^d : \quad \langle u_i, x \rangle \leq 0 \quad \text{for} \quad i = 1, \ldots, d \Big\}$$

for some linearly independent vectors $u_1, \ldots, u_d \in \mathbb{Z}^d$. Prove that K is a simple rational cone.

7. Let $u_1, \ldots, u_d \in \mathbb{Z}^d$ be linearly independent vectors and let $u_1^*, \ldots, u_d^*$ be defined by $\langle u_i, u_j \rangle = -\delta_{ij}$. Prove that $u_1^*, \ldots, u_d^* \in \mathbb{Z}^d$ if and only if the cone K generated by $u_1, \ldots, u_d$ is unimodular (we assume that $u_1, \ldots, u_d$ are the minimal generators of K).

Supplementary problems.

1. Let $u_1, \ldots, u_d$ be linearly independent vectors in $\mathbb{Z}^d$. Let K be the cone generated by $u_1, \ldots, u_d$ and let

$$\operatorname{int} K = \Big\{ \sum_{i=1}^d \alpha_i u_i : \quad \alpha_i > 0 \quad \text{for} \quad i = 1, \ldots, d \Big\}$$

be the interior of K. Let

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

Prove that

$$f(\operatorname{int} K, \mathbf{x}) = \left(\sum_{m \in \overline{\Pi} \cap \mathbb{Z}^d} \mathbf{x}^m \right) \prod_{i=1}^d \frac{1}{1 - \mathbf{x}^{u_i}}.$$

Deduce the *reciprocity relation* $f(\operatorname{int} K, \mathbf{x}^{-1}) = (-1)^d f(K, \mathbf{x})$.

2. Deduce from Theorem 2 the following *Pick's* Theorem: if $P \subset \mathbb{R}^2$ is a convex polygon with integer vertices and non-empty interior, then the number of integer points in P is equal to the area of P plus half of the number of integer points on the boundary of P plus 1, see Figure 15.

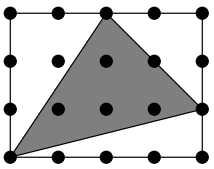

Figure 15. The number of integer points in the triangle (8) is equal to the area of the triangle (5) plus half of the number of integer points on the boundary (2) plus 1

Preview problems.

1. Let $K \subset \mathbb{R}^d$ be a unimodular cone generated by integer vectors $u_1, \ldots, u_d$ and let $K + v$ be the translation of K by a rational vector $v \in \mathbb{Q}^d$. Prove that

$$f(K + v, \mathbf{x}) = \mathbf{x}^w \prod_{i=1}^d \frac{1}{1 - \mathbf{x}^{u_i}} \quad \text{with} \quad w = \sum_{i=1}^d \lceil \langle v, u_i^* \rangle \rceil u_i,$$

where $u_1^*, \ldots, u_d^*$ are defined by $\langle u_i^*, u_j \rangle = \delta_{ij}$.

2. Construct an efficient (polynomial time) algorithm to *sample* a random integer point in a given fundamental parallelepiped Π from the uniform distribution on $\Pi \cap \mathbb{Z}^d$ (the dimension d needs not to be fixed in advance).

Remarks

For generating functions and rational cones, see Section 4.6 of [**St97**] and Section VIII.1 of [**Ba02**]. A classical reference for continued fractions is [**Kh97**]. For the theory of computational complexity, see [**Pa94**].

LECTURE 4
Rational Polyhedra and Rational Functions

Let $P \subset \mathbb{R}^d$ be a rational polyhedron. Our goal is to understand the generating function

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

Previously, we discussed what happens if $P = K$ is a simple rational cone. Step by step, we go to larger and larger classes of polyhedra.

Definition 1. A rational polyhedron $K \subset \mathbb{R}^d$ is called a *rational cone* provided $0 \in K$ and $\lambda x \in K$ for every $x \in K$ and every $\lambda \geq 0$. Equivalently, K is a rational cone if K can be defined by a system of finitely many homogeneous linear inequalities with integer coefficients:

$$K = \Big\{ x : \quad \sum_{j=1}^{d} a_{ij} x_j \leq 0 \quad \text{for} \quad i = 1, \ldots, m \Big\}.$$

If 0 is a vertex of K, the cone is called *pointed*.

The first real difference between the concept of a rational cone and that of a simple rational cone transpires at $d = 3$. While simple rational cones in $\mathbb{R}^d$ are defined by exactly d inequalities, non-simple rational cones may require typically more or sometimes fewer inequalities.

Theorem 1. *Let $K \subset \mathbb{R}^d$ be a pointed rational cone. Then $f(K, \mathbf{x})$ is a rational function in $\mathbf{x}$ of the type*

$$f(K, \mathbf{x}) = \sum_{i=1}^{n} \frac{p_i(\mathbf{x})}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $p_i(\mathbf{x})$ are Laurent polynomials in $\mathbf{x}$ and $u_{ij} \in \mathbb{Z}^d$ are non-zero vectors.

A plausible argument. Since 0 is a vertex of K, there is a vector $c \in \mathbb{R}^d$, $c = (c_1, \ldots, c_d)$ such that $\langle c, x \rangle < 0$ for all $x \in K \setminus \{0\}$. Now, for any $\mathbf{x}$ from a sufficiently small neighborhood U of $x_0 = (e^{c_1}, \ldots, e^{c_d})$ the series $\sum_{m \in K \cap \mathbb{Z}^d} \mathbf{x}^m$ converges absolutely and uniformly on compact subsets of U. It seems intuitively obvious and indeed correct that K can be cut into simple rational cones, so we

can deduce the formula for $f(K, \mathbf{x})$ from Theorem 1, Lecture 3, and the inclusion-exclusion formula. It takes some time though to make the proof rigorous, cf. Figure 16. □

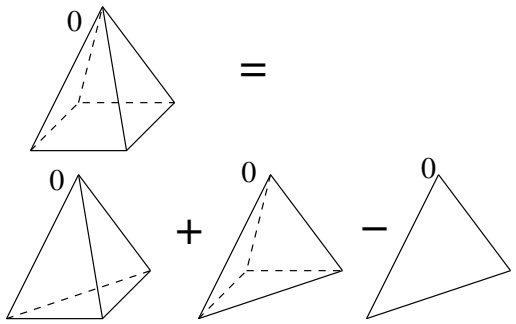

Figure 16. Example: The indicator of a cone with a square base can be written as the sum of the indicators of cones with triangular bases minus the indicator of a flat cone based on the interval

Next, we consider an arbitrary rational polyhedron with a vertex.

Theorem 2. *Let $P \subset \mathbb{R}^d$ be a rational polyhedron with a vertex (equivalently, a non-empty rational polyhedron without lines). Then $f(P, \mathbf{x})$ is a rational function of the type*

$$f(P, \mathbf{x}) = \sum_{i=1}^{n} \frac{p_i(\mathbf{x})}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $p_i(\mathbf{x})$ are Laurent polynomials in $\mathbf{x}$ and $u_{ij} \in \mathbb{Z}^d$ are non-zero vectors.

Sketch of proof. The idea is to consider P as a section of a pointed rational cone $K \subset \mathbb{R}^{d+1}$. We think of $\mathbb{R}^d$ as the affine hyperplane $x_{d+1} = 1$ in $\mathbb{R}^{d+1}$. Given the inequalities defining P

$$\sum_{j=1}^{d} a_{ij} x_j \leq b_i \quad \text{for} \quad i = 1, \ldots, m,$$

we define K by the inequalities

$$\sum_{j=1}^{d} a_{ij} x_j - b_i x_{d+1} \leq 0, \quad x_{d+1} \geq 0.$$

Clearly, K is a rational cone and P is the section of K by the flat $x_{d+1} = 1$, cf. Figure 17.

One can also prove that K is pointed via the following chain of implications:

P contains no lines $\implies$ K contains no lines $\implies$ K is pointed.

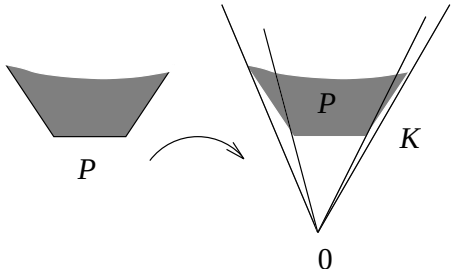

Figure 17. Representing a d-dimensional polyhedron P as a hyperplane section of a $(d+1)$-dimensional cone K

Finally, we obtain $f(P, \mathbf{x})$ by differentiating with respect to x_{d+1}:

$$f(P, \mathbf{x}) = \frac{\partial}{\partial x_{d+1}} f(K, \mathbf{x}) \quad \text{evaluated at} \quad x_{d+1} = 0.$$

□

Suppose now that P is a rational polyhedron with lines. In this case, the series $\sum_{m \in P \cap \mathbb{Z}^d} \mathbf{x}^m$ does not converge anywhere. As we hinted in the introductory examples of Lecture 1, we want to define $f(P, \mathbf{x}) \equiv 0$ in this case. Quite surprisingly, this naive solution works just fine. The following remarkable result was proved by J. Lawrence, and, independently, by A. Khovanski and A. Pukhlikov in early 1990's.

Theorem 3. *There exists a map*

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

from the algebra of rational polyhedra to the ring of rational functions in d variables $\mathbf{x} = (x_1, \ldots, x_d)$ such that

1. *The map $\mathcal{F}$ is a valuation, that is, a linear transformation,*
2. *If $P \subset \mathbb{R}^d$ is a rational polyhedron without lines then $\mathcal{F}[P] = f(P, \mathbf{x})$ is the rational function such that*

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

   provided the series converges absolutely;
3. *If $P \subset \mathbb{R}^d$ is a polyhedron containing a line then $\mathcal{F}[P] = 0$.*

Sketch of proof. We know how to define $\mathcal{F}$ on the indicators $[P]$ of rational polyhedra P without lines, as in Part (2) of the theorem. Our proof consists of two steps:

the first step is to show that $\mathcal{F}$ can be extended to a valuation on $\mathcal{P}(\mathbb{Q}^d)$;

the second step is to show that once we extended $\mathcal{F}$ to a valuation, we necessarily have $\mathcal{F}[P] = 0$ for rational polyhedra P with lines.

It is clear how we *should* extend $\mathcal{F}$ onto polyhedra with lines (it is not yet clear that we *can*). Any rational polyhedron P can be cut into rational polyhedral pieces P_i without lines, so we *should* compute $\mathcal{F}[P]$ from $\mathcal{F}[P_i] = f(P_i, \mathbf{x})$ via the inclusion-exclusion formula. The problem is, of course, to show that this extension is not self-contradictory. This, in turn, reduces to proving that whenever we have a linear dependence

$$\sum_{i=1}^{n} \alpha_i [P_i] = 0 \tag{1}$$

of indicators of rational polyhedra P_i without lines, we must have the same linear dependence

$$\sum_{i=1}^{n} \alpha_i f(P_i, \mathbf{x}) = 0 \tag{2}$$

of their generating functions. Suppose for a moment that in (1) there exists a non-empty open set $U \subset \mathbb{C}^d$ such that for $\mathbf{x} \in U$ *all* the series $\sum_{m \in P_i \cap \mathbb{Z}^d} \mathbf{x}^m$ converge absolutely to $f(P_i, \mathbf{x})$. Then (2) follows by a standard argument from analysis. The problem is that such an U, same for all polyhedra P_i in (1), might not exist. To handle this difficulty, we break the global identity (1) into small "local" pieces, prove (2) for every such piece and then "glue" the global identity (2) from the local pieces.

For a non-empty subset $I \subset \{1, \ldots, n\}$, let

$$P_I = \bigcap_{i \in I} P_i.$$

Clearly, P_I are rational polyhedra without lines, possibly empty.

From the inclusion-exclusion formula, we have

$$\left[\bigcup_{i=1}^{n} P_i \right] = \sum_{I} (-1)^{|I|-1} \left[P_I \right].$$

Multiplying both sides by $[P_i]$, we get the formula

$$[P_i] = \sum_{I} (-1)^{|I|-1} \left[P_{I \cup \{i\}} \right].$$

The crucial observation is that P_i is a rational polyhedron without lines and that $P_{I \cup \{i\}}$ are rational polyhedral pieces of P_i. Therefore, there is a non-empty open set $U \subset \mathbb{C}^d$ such that for all $\mathbf{x} \in U$ all the series defining $f(P_i, \mathbf{x})$ and $f(P_{I \cup \{i\}}, \mathbf{x})$ converge and so we have the identity

$$f(P_i, \mathbf{x}) = \sum_{I} (-1)^{|I|-1} f(P_{I \cup \{i\}}, \mathbf{x}). \tag{3}$$

Next, multiplying (1) by $[P_I]$, we get

$$\sum_{i=1}^{n} \alpha_i [P_{I\cup\{i\}}] = 0.$$

Again, all $P_{I\cup\{i\}}$ are rational polyhedral pieces of a rational polyhedron P_i without lines, and, since we can find a single domain $U \subset \mathbb{C}^d$ for which all the relevant series converge, we get

$$\sum_{i=1}^{n} \alpha_i f(P_{I\cup\{i\}}, \mathbf{x}) = 0. \tag{4}$$

From (4) and (3) we get (2). This completes the first step of the proof.

Thus we are able to extend $\mathcal{F}$ to a valuation on $\mathcal{P}(\mathbb{Q}^d)$. It remains to prove Part (3) of the Theorem. One can show that if P is a rational polyhedron with lines, then there exists a non-zero $m \in \mathbb{Z}^d$ such that $P + m = P$ (there is a non-zero integer translation of P which maps P onto itself). On the other hand, from elementary analysis we deduce that we must have $f(P+m, \mathbf{x}) = \mathbf{x}^m f(P, \mathbf{x})$ for any rational polyhedron P without lines. By linearity, $\mathcal{F}[P + m] = \mathbf{x}^m \mathcal{F}[P]$ for *any* rational polyhedron P. Hence, if $P + m = P$, we must have $\mathcal{F}[P] = \mathbf{x}^m \mathcal{F}[P]$, from which $\mathcal{F}[P] = 0$. □

Suppose that $P \subset \mathbb{R}^d$ is a rational polyhedron without lines (maybe even bounded) and that we want to compute a short formula for the rational generating function $f(P, \mathbf{x})$. Theorem 3 allows us to employ various identities in the algebra $\mathcal{P}(\mathbb{Q}^d)$ of rational polyhedra, including those that involve polyhedra with lines. In particular, we get the following result, first obtained by M. Brion in 1988.

Theorem 4. *Let $P \subset \mathbb{R}^d$ be a rational polyhedron with vertices. Then*

$$f(P, \mathbf{x}) = \sum_{v} f\big(\mathrm{co}(P, v), \mathbf{x}\big),$$

where the sum is taken over all vertices v of P and $\mathrm{co}(P, v)$ is the tangent cone of P at v.

Proof. The proof follows by Theorem 3 of this lecture and Theorem 3 of Lecture 2. □

Note that the tangent cone $\mathrm{co}(P, v)$ is not a rational cone per se, but a rational translation of a rational cone.

Example 1. Let $d = 1$ and let P be the interval $[0, n] \subset \mathbb{R}^1$ for some positive integer n. Then P is a rational polyhedron with the vertices at 0 and n, see Figure 18. The tangent cone $\mathrm{co}(P, 0)$ at 0 is the ray $[0, +\infty)$ and the corresponding generating function is

$$\sum_{m=0}^{+\infty} x^m = \frac{1}{1-x}.$$

The tangent cone $\mathrm{co}(P, n)$ at n is the ray $(-\infty, n]$ and the corresponding generating function is

$$\sum_{m=-\infty}^{n} x^m = \frac{x^n}{1-x^{-1}} = \frac{-x^{n+1}}{1-x}.$$

Note that there is not a single value of x for which both series converge. Nevertheless, Theorem 4 predicts that the sum of the two functions gives the generating function for P:

$$\sum_{m=0}^{n} x^m = \frac{1}{1-x} - \frac{x^{n+1}}{1-x},$$

which is indeed the case.

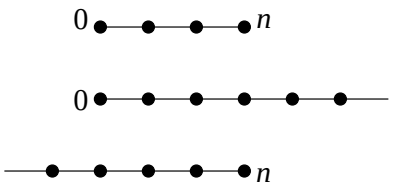

Figure 18. An interval and its tangent cones

Problems

Review problems.

1. Complete the proof of Theorem 2.

2. Let $P \subset \mathbb{R}^d$ be a rational polyhedron with a line. Prove that there exists a non-zero vector $m \in \mathbb{Z}^d$ such that $P + m = P$.

3. Complete the proof of Theorem 3.

4. Check Theorem 4 for the triangle in the plane with the vertices $(0,0)$, $(0,1)$, and $(1,0)$.

A supplementary problem.

1. Let $K \subset \mathbb{R}^d$ be a pointed rational cone with non-empty interior $\operatorname{int} K$. Prove the *reciprocity relation* $f(\operatorname{int} K, \mathbf{x}^{-1}) = (-1)^d f(K, \mathbf{x})$.

Preview problems

1. Prove that the polar of a unimodular cone is a unimodular cone.

2. Let $K \subset \mathbb{R}^2$ be the cone generated by $u_1 = (1,0)$ and $u_2 = (q,p)$ some positive integer p and q. Compare the following two ways of computing $f(K, \mathbf{x})$. The first way is the continued fractions method of Lecture 3. The second way is as follows: consider the polar K° (check that K° is the cone generated by $(-p,q)$ and $(0,-1)$). Represent $[K^\circ]$ as a linear combination of the indicators of unimodular cones using the continued fractions method. Apply Theorem 4 of Lecture 1 to obtain a unimodular decomposition of $K = (K^\circ)^\circ$. Compute $f(K, \mathbf{x})$ from that decomposition. What kind of identities do we get for $f(K, \mathbf{x})$?

This question was asked by one of the attendees.

Remarks

Theorem 3 is proved in [**La91**] and, independently, in [**KP92**]. The first proof of Theorem 4 [**Br88**] uses algebraic geometry. For the material of this lecture, see Sections VIII.3–4 of [**Ba02**] (Theorems 3 and 4) and Section 4.6 of [**St97**] (generating functions for rational cones and the reciprocity relation).

LECTURE 5
Computing Generating Functions Fast

We discuss *how* to compute the generating function $f(P, \mathbf{x})$ fast, but before we do that, we discuss *why* we want to compute it and what *fast* means.

Why do we need generating functions

Let $P \subset \mathbb{R}^d$ be a bounded rational polyhedron (rational polytope). For a variety of reasons, we need to compute the number $|P \cap \mathbb{Z}^d|$ of integer points in P (the counting problem). If we are able to compute the generating function

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

which is just a Laurent polynomial in $\mathbf{x}$, we can get the number of integer points $|P \cap \mathbb{Z}^d|$ by substituting $\mathbf{x} = (1, \dots, 1)$. Our technique allows us to compute $f(P, \mathbf{x})$ as a reasonably short rational function of the type

$$f(P, \mathbf{x}) = \sum_i \frac{p_i(\mathbf{x})}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $p_i(\mathbf{x})$ are Laurent polynomials in $\mathbf{x}$. This seems to pose a little problem since $\mathbf{x} = (1, \dots, 1)$ is a pole of every fraction. Nevertheless, the poles cancel each other, as in the model example

$$\sum_{m=0}^{n} x^m = \frac{1}{1-x} - \frac{x^{n+1}}{1-x}.$$

We deal with singularities by approaching the point $(1, \dots, 1)$ via some curve and computing the appropriate limit. One of the standard choices is the curve $\mathbf{x}(t) = (e^{tc_1}, \dots, e^{tc_d})$, where $c = (c_1, \dots, c_d)$ is a sufficiently generic vector: we need $\langle c, u_{ij} \rangle \neq 0$ for all i, j. Then $\mathbf{x}(0) = (1, \dots, 1)$ and the limit $f(P, \mathbf{x}(t))$ as $t \longrightarrow 0$ can be computed by using the standard analysis technique.

Generating functions help to solve *integer programming problems*, that is the problems of optimizing a given linear function on the set $P \cap \mathbb{Z}^d$ of integer points in a given rational polytope. In short, generating functions $f(P, \mathbf{x})$ encode all the information about the set of integer points in P in a compact form. One remarkable fact is that to find a short formula for $f(P, \mathbf{x})$ for a *bounded* polyhedron, we employ the full power of the algebra $\mathcal{P}(\mathbb{Q}^d)$ and identities in the algebra involving *unbounded* polyhedra and even polyhedra with lines (Theorems 3 and 4 of Lecture 4).

What "fast" and "short" means

We mentioned more than once that we want to compute the generating function $f(P, \mathbf{x})$ "fast" and that we want it in a "reasonably short" form. The exact meaning of these words is understood through the theory of computational complexity.

The polyhedron P is defined by a system of linear inequalities

$$\sum_{j=1}^{d} a_{ij} x_j \leq b_i, \quad i = 1, \dots, n.$$

The *input size* of P is the number of bits needed to write down the inequalities, assuming that a_{ij} and b_i are integers written in the binary system. For example, to write an integer a, we need about $\log |a| + O(1)$ bits. Thus we say that the algorithm for computing $f(P, \mathbf{x})$ is reasonably fast and the resulting formula is reasonably short if the time we need to compute $f(P, \mathbf{x})$ and the space we need to write it down grows only modestly when the input size of P grows. More precisely, we say that we have a *polynomial time* algorithm for a particular class of rational polyhedra if there is a polynomial *poly* such that the running time of the algorithm on every polyhedron P from the class does not exceed *poly*(input size of P). One example of a polynomial time algorithm is provided by the continued fraction method for computing $f(K, \mathbf{x})$ where K is a 2-dimensional rational cone, see Lecture 3.

It is probably hopeless to search for a polynomial time algorithm in the class of *all* rational polyhedra. However, once the dimension d is fixed such algorithms exist.

Theorem 1. *Let us fix d. Then there exists a polynomial time algorithm, which, given a rational polyhedron $P \subset \mathbb{R}^d$, computes the generating function $f(P, \mathbf{x})$ in the form*

$$f(P, \mathbf{x}) = \sum_i \alpha_i \frac{\mathbf{x}^{v_i}}{(1 - \mathbf{x}^{u_{i1}}) \cdots (1 - \mathbf{x}^{u_{id}})},$$

where $\alpha_i \in \{-1, 1\}$, $v_i \in \mathbb{Z}^d$, and $u_{ij} \in \mathbb{Z}^d \setminus \{0\}$ for all i, j.

Since the running time of the algorithm includes the time needed to write down the output, the space needed to write down $f(P, \mathbf{x})$ is also bounded by a polynomial in the input size.

There exist several versions of the main algorithm behind Theorem 1. Different versions have different advantages under different circumstances. Moreover, the algorithm of Theorem 1 appears to be *practical*. It has been implemented (in fact, by at least two independent groups). The main procedure behind Theorem 1 is a certain *unimodular cone decomposition*. We sketch it below.

Preliminaries

The main result we need is Minkowski's *Convex Body Theorem*. Let $A \subset \mathbb{R}^d$ be a set, such that

A is convex, that is, for every two points $x, y \in A$, the interval $[x, y] = \big\{\alpha x + (1 - \alpha) y : \ 0 \leq \alpha \leq 1\big\}$ also lies in A;

A is symmetric about the origin, that is, for every $x \in A$, the point $-x$ also lies in A;

A has a sufficiently large volume: $\operatorname{vol} A > 2^d$.

Moreover, if A is compact, we may assume that $\operatorname{vol} A \geq 2^d$.

Minkowski's Convex Body Theorem asserts that A necessarily contains a non–zero integer point. Here is the idea of the proof: consider $X = \{x/2 : \ x \in A\}$, so that $\operatorname{vol} X > 1$. Consider the set of all integer translates $X + u : \ u \in \mathbb{Z}^d$. Argue that some two different translates must overlap: $(X + u_1) \cap (X + u_2) \neq \emptyset$. Deduce that there is a non-zero lattice point in A.

If A is a rational polyhedron, then such a non-zero integer point in A can be found efficiently, though we don't discuss how.

The unimodular decomposition of a cone

Let $K \subset \mathbb{R}^d$ be a simple rational cone generated by linearly independent vectors $u_1, \ldots, u_d \in \mathbb{Z}^d$. Our goal is to construct unimodular cones K_i such that

$$[K] = \sum_i \alpha_i [K_i] \quad + \quad \text{indicators of lower-dimensional cones}$$

and $\alpha_i \in \{-1, 1\}$. The algorithm runs in polynomial time if the dimension d fixed.

Let Π be the fundamental parallelepiped of K. Then $\operatorname{vol} \Pi$ is a positive integer and $\operatorname{vol} \Pi = 1$ if and only if K is unimodular. Let us call $\operatorname{vol} \Pi$ the *index* of K and denote it $\operatorname{ind} K$. Thus $\operatorname{ind} K$ measures how far is K from being unimodular. We will iterate a certain procedure which gradually reduces the index of K.

Let

$$A = \Big\{ \sum_{i=1}^{d} \alpha_i u_i : \quad |\alpha_i| \leq (\operatorname{ind} K)^{-1/d} \Big\}.$$

Then $\operatorname{vol} A = 2^d$ and by Minkowski's Convex Body Theorem there is a non-zero point $v \in A \cap \mathbb{Z}^d$, cf. Figure 19.

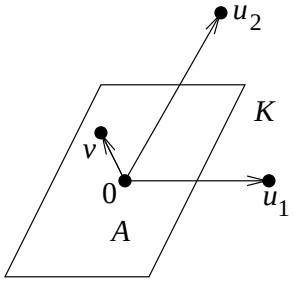

Figure 19. Finding the point

As we mentioned, we can find this point efficiently if the dimension d is fixed.

For $i = 1, \ldots, d$, let K_i be the cone generated by $u_1, \ldots, u_{i-1}, v, u_{i+1}, \ldots, u_d$. Then

$$\operatorname{ind} K_i \leq (\operatorname{ind} K)^{\frac{d-1}{d}} .$$

Let $\alpha_i = 1$ or $\alpha_i = -1$ depending on whether the orientations of the bases $u_1, \dots, u_{i-1}, v, u_{i+1}, \dots, u_d$ and $u_1, \dots, u_d$ are the same or the opposite. Then

$$[K] = \sum_{i=1}^{d} \alpha_i [K_i] \quad + \quad \text{indicators of lower-dimensional cones,}$$

see Figure 20.

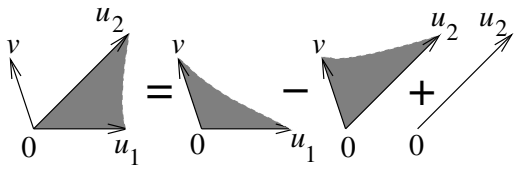

Figure 20. Writing the cone as a linear combination of cones with smaller indices

Now we iterate the procedure. After k iterations, we get d^k cones K_i with

$$\operatorname{ind} K_i \le (\operatorname{ind} K)^{\left(\frac{d-1}{d}\right)^k}.$$

In other words, the number of cones grows exponentially in the number k of iterations while the indices of the cones decrease doubly exponentially in k. It follows that when d is fixed, to obtain a unimodular decomposition, we need $k = O(\log\log(\operatorname{ind} K))$ iterations, which results in a polynomial time algorithm.

There are certain similarities between the described procedure and the unimodular decomposition obtained from the continued fractions method in dimension 2. There are differences, too. In the method just described, there is a certain flexibility in choosing vector v, while the continued fractions method is quite rigid. This is, of course, due to the fact that we know much more about integer points in dimension 2 than in higher dimensions. On the other hand, there is a version of our algorithm that reduces to the continued fractions method in dimension 2.

Polarity and discarding lower-dimensional cones

When we apply the procedure described above, there is an apparent nuisance of dealing with lower-dimensional cones. However, there is a certain trick which allows us simply to forget about them.

Let $K \subset \mathbb{R}^d$ be a simple rational cone. Let

$$K^\circ = \left\{ x \in \mathbb{R}^d : \quad \langle x, y \rangle \le 0 \quad \text{for all} \quad y \in K \right\}$$

be the polar of K, see Lecture 2. It is not hard to prove that K° is a simple rational cone, that K° is unimodular if and only if K is unimodular, and that $(K^\circ)^\circ = K$. Thus we modify the above procedure as follows.

Given a simple rational cone K, we compute the polar K°. Then we apply the unimodular decomposition and get

$$[K^\circ] = \sum_i \alpha_i [K_i] \quad + \quad \text{indicators of lower-dimensional cones,}$$

where K_i are unimodular cones. Next, we compute K_i° and observe that

$$[K] = \sum_i \alpha_i [K_i^\circ] \quad + \quad \text{indicators of cones with lines,}$$

see Theorem 4 and Review Problem 14 of Lecture 1.

By Theorem 3 of Lecture 4,

$$f(K, \mathbf{x}) = \sum_i \alpha_i f(K_i^\circ, \mathbf{x}),$$

since we can ignore polyhedra with lines.

Problems

Review problems

1. Check that the procedure of computing $f(K, \mathbf{x})$ for a simple rational cone $K \subset \mathbb{R}^2$ via continued fractions (see Lecture 3) indeed runs in polynomial time.

2. Prove Minkowski's Convex Body Theorem.

3. Check that the algorithm for the unimodular decomposition of a cone indeed works.

4. Let $K \subset \mathbb{R}^d$ be a unimodular cone. Prove that K° is a unimodular cone.

Supplementary problems.

1. Let a_1 and a_2 be positive coprime integers and let $S \subset \mathbb{Z}$ be the set of all non-negative integer combinations of a_1 and a_2. Prove that

$$\sum_{m \in S} x^m = \frac{1 - x^{a_1 a_2}}{(1 - x^{a_1})(1 - x^{a_2})}.$$

2. Let a_1, a_2 and a_3 be positive comprime integers and let $S \subset \mathbb{Z}$ be the set of all non-negative integer combinations of a_1, a_2 and a_3. Prove that

$$\sum_{m \in S} = \frac{1 - x^{b_1} - x^{b_2} - x^{b_3} + x^{b_4} + x^{b_5}}{(1 - x^{a_1})(1 - x^{a_2})(1 - x^{a_3})},$$

for some, not necessarily distinct, integers $b_i = b_i(a_1, a_2, a_3)$, $i = 1, 2, 3, 4, 5$.

3. Let $a_1, \ldots, a_n \in \mathbb{Z}_+^d$ be some integer vectors with non-negative coordinates. Let $S \subset \mathbb{Z}_+^d$ be the set of all non-negative integer combinations of $a_1, \ldots, a_n$. Prove that the series

$$\sum_{m \in S} \mathbf{x}^m \quad \text{where} \quad \mathbf{x} = (x_1, \ldots, x_d)$$

converges for $|x_i| < 1$, $i = 1, \ldots, d$, to a rational function of $\mathbf{x}$.

Concluding remarks

The algorithmic theory of counting lattice points in polyhedra is discussed in [**BP99**]; some of the algorithms suggested there are implemented, see [**L+04**] and [**V+04**]. For other algorithmic questions concerning lattice points, see [**G+93**]. For Minkowski's Theorems and other topics in the geometry of numbers, see [**GL87**].

We conclude these lectures by discussing various related topics and open questions.

Something curvilinear? Is it possible to extend the developed theory onto something non-polyhedral, such as Euclidean balls? Probably not, as it appears to be in the realm of totally different forces, more akin to theta functions than to rational functions. For example, let

$$B = \left\{ (x_1, \ldots, x_4) : \quad \sum_{i=1}^{4} x_i^2 \leq n \right\}$$

be the standard Euclidean ball of radius $\sqrt{n}$ in dimension 4. Suppose for a moment that we can efficiently enumerate integer points in B. Then we can count integer points on the sphere $x_1^2 + x_2^2 + x_3^2 + x_4^2 = n$. However, the number of such points, that is, the number of ways to represent n as a sum of four squares of integers, by Jacobi's formula is equal to

$$8 \sum_{4 \nmid p \mid n} p$$

(in words: eight times the sum of the divisors of n that are not divisible by 4). Thus we gain some insight into divisors of n, and, pushing it a bit further, we can come up with an efficient algorithm for factoring integers, see [**B+86**] and [**Dy91**]. The existence of such an algorithm is not entirely impossible, but somewhat doubtful.

Irrational polyhedra? How can we enumerate integer points in irrational polyhedra? There are some obvious difficulties with generating functions. Consider, for example, a cone $K \subset \mathbb{R}^2$ defined by the inequalities $x_1 \geq 0$ and $x_2 \leq \sqrt{2} x_1$. Just as before, we can write the generating function

$$f(K, \mathbf{x}) = \sum_{m \in K \cap \mathbb{Z}^2} \mathbf{x}^m.$$

The problem is that $f(K, \mathbf{x})$ is no longer a rational function in $\mathbf{x}$. To build an interesting theory, we would like to extend $f(K, \mathbf{x})$ analytically far beyond the region of convergence of the defining series, and it is not clear how to do that.

There is a little trick, however, which allows us to incorporate irrational polyhedra P to some extent. Let us first change the coordinates and consider the *exponential sum*:

$$F(P; c) = \sum_{m \in P \cap \mathbb{Z}^d} e^{\langle c, m \rangle},$$

where $c = (c_1, \ldots, c_d) \in \mathbb{C}^d$. We obtain $f(P, \mathbf{x})$ by substituting $\mathbf{x} = (e^{c_1}, \ldots, e^{c_d})$. Let $\rho : \mathbb{C}^d \longrightarrow \mathbb{C}$ be a polynomial and let us consider the weighted version

$$F(P, \rho; c) = \sum_{m \in P \cap \mathbb{Z}^d} \rho(m) e^{\langle c, m \rangle}$$

of the exponential sum. One can think of $F(P, \rho; c)$ as the result of applying the differential operator

$$D = \rho\left(\frac{\partial}{\partial c_1}, \ldots, \frac{\partial}{\partial c_d}\right)$$

to $F(P; c)$. If P is an irrational polyhedron, all "bad things" happen along the boundary ∂P of P, so let us cut them out by choosing ρ such that $\rho(x) = 0$ for all $x \in \partial P$ (such a ρ can be obtained by multiplying the equations that define the facets of P). One can show that in this case $F(P, \rho; c)$ indeed extends to a meromorphic function on $\mathbb{C}^d$ and there is a way to extend our theory, see [**Ba93**] for details. This extension, however, is not particularly interesting (it lacks interesting examples so far).

Let's add projections! There are interesting sets $S \subset \mathbb{Z}^d$ of integer points, which are intimately related to sets of integer points in rational polyhedra but have a more complicated logical structure. Such sets can be quite complicated even in dimension $d = 1$. For example, let us fix positive coprime integers $a_1, \ldots, a_d$ and let $S \subset \mathbb{Z}$ be the set of all integers that are non-negative integer combinations of $a_1, \ldots, a_d$. In other words, S is the semigroup generated by $a_1, \ldots, a_d$. We can think of S as a *projection* of the set of integer points in a rational polyhedron. Let $P = \mathbb{R}^d_+$ be the non-negative orthant and let $T : \mathbb{R}^d \longrightarrow \mathbb{R}$ be the projection

$$(x_1, \ldots, x_d) \longmapsto a_1 x_1 + \ldots + a_d x_d.$$

Then $S = T(P \cap \mathbb{Z}^d)$, the image of the set of integer points in P under the linear transformation T. In [**BW**], it is proved that for such sets S (obtained from the set of integer points $P \cap \mathbb{Z}^d$ in a rational polyhedron $P \subset \mathbb{R}^d$ by a projection) the generating function

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

admits a short representation as a rational function in $\mathbf{x}$, which can be computed in polynomial time when the dimension d is fixed.

There are some advances towards the general theory of sets of integer points encoded by short rational generating functions. For example, in [**BW**] it is proved that if two sets $S_1, S_2 \subset \mathbb{Z}^d$ are *defined* by their short rational generating functions $f(S_1, \mathbf{x})$ and $f(S_2, \mathbf{x})$, then the generating function $f(S, \mathbf{x})$ of $S = S_1 \cap S_2$ can be computed in polynomial time as a short rational function.

However, we are still quite far from having a full-fledged theory for sets S with short rational generating functions. Suppose, for example, that S is the projection of the difference $X \setminus Y$, where X and Y are the projections of the sets of integer points $P \cap \mathbb{Z}^{k_1}$ and $Q \cap \mathbb{Z}^{k_2}$ in some rational polyhedra P and Q. We don't know how to handle such a set S (our lack of understanding is mitigated by the lack of interesting examples of such complicated constructions). Also, algorithms of [**BW**] seem to be outrageously impractical.

Polytopes of large dimension? The theory we described in these lectures provides efficient algorithms if the dimension d of the given rational polytope is fixed in advance. There are certain classes of polytopes of large (that is, allowed to grow) dimension d for which the algorithms still remain efficient (polynomial time), see [**Ba93**] and [**BP99**]. However, if the dimebsion d is allowed to grow, because of the

P vs. **NP** issue, one cannot hope to test efficiently whether a given rational polyhedron contains an integer point, let alone to compute the number of such points efficiently. The problem, however, remains practically important and it seems that various probabilistic approaches of approximate counting look the most promising here, cf. [**Dy03**].

There seems to be a possibility of a "hybrid" algebraic/probabilistic approach based on the following simple observation. Let $K = K(u_1, \ldots, u_d)$ be a simple rational cone generated by integer vectors $u_1, \ldots, u_d$ and let Π be its fundamental parallelepiped. Theorem 1 of Lecture 3 allows us to compute $f(K, \mathbf{x})$ in terms of the sum $\sum_{m \in \Pi \cap \mathbb{Z}^d} \mathbf{x}^m$. This sum is potentially very big, but it is very easy to *sample* a random integer point $m \in \Pi \cap \mathbb{Z}^d$, cf. Preview Problem 2 in Lecture 3. Indeed, let $\Lambda \subset \mathbb{Z}^d$ be the lattice generated by $u_1, \ldots, u_d$. Then the points $\Pi \cap \mathbb{Z}^d$ are in one-to-one correspondence with the elements of $\mathbb{Z}^d/\Lambda$: if $n \in \mathbb{Z}^d$ is an integer point, then the point $m \in \Pi \cap \mathbb{Z}^d$ such that $m - n \in \Lambda$ is computed as follows: we write $n = \sum_{i=1}^d \alpha_i u_i$ and let $m = \sum_{i=1}^d \{\alpha_i\} u_i$. Hence the problem of sampling $m \in \Pi \cap \mathbb{Z}^d$ reduces to that of sampling coset representatives $n \in \mathbb{Z}^d/\Lambda$, which can be done efficiently. One can ask what happens if we try to count integer points in a given polytope P by using Brion's Theorem (Theorem 4 of Lecture 4) and computing the generating functions of the supporting cones of P approximately via random sampling?

BIBLIOGRAPHY

[Ba93] A. Barvinok, *Computing the volume, counting integral points, and exponential sums*, Discrete Comput. Geom. **10** (1993), 123–141.

[Ba02] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics, vol. 54, Amer. Math. Soc., Providence, RI, 2002.

[Br88] M. Brion, *Points entiers dans les polyédres convexes (French)*, Ann. Sci. École Norm. Sup. (4) **21** (1988), 653–663.

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

[BW03] A. Barvinok and K. Woods, *Short rational generating functions for lattice point problems*, J. Amer. Math. Soc. **16** (2003), 957–979.

[B+86] E. Bach, G. Miller, and J. Shallit, *Sums of divisors, perfect numbers and factoring*, SIAM J. Comput. **15** (1986), 1143–1154.

[Dy91] M. Dyer, *On counting lattice points in polyhedra*, SIAM J. Comput. **20** (1991), 695–707.

[Dy03] M. Dyer, *Approximate counting by dynamic programming*, Proceedings of the 35th Annual ACM Symposium on the Theory of Computing (STOC 2003), 2003, pp. 693–699.

[GL87] P.M. Gruber and C.G. Lekkerkerker, *Geometry of Numbers. Second edition*, North-Holland Mathematical Library, vol. 37, North-Holland, Amsterdam, 1987.

[G+93] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization. Second edition*, Algorithms and Combinatorics, vol. 2, Springer-Verlag, Berlin, 1993.

[Kh97] A. Ya. Khinchin, *Continued Fractions*, Translated from the third (1961) Russian edition. Reprint of the 1964 translation, Dover Publications, Inc., Mineola, NY, 1997.

[KR97] D. Klain and G.-C. Rota, *Introduction to geometric probability*, Lezioni Lincee, Cambridge Univ. Press, Cambridge, 1997.

[KP92] A.G. Khovanskii and A.V. Pukhlikov, *The Riemann-Roch theorem for integrals and sums of quasipolynomials on virtual polytopes. (Russian)*, translation in St. Petersburg Math. J. 4 (1993), no. 4, 789–812, Algebra i Analiz **4, no. 4** (1992), 188–216.

[KP93] A.G. Khovanskii and A.V. Pukhlikov, *Integral transforms based on Euler characteristic and their applications*, Integral Transform. Spec. Funct. **1** (1993), 19–26.

[La91] J. Lawrence, *Rational-function-valued valuations on polyhedra*, Discrete and computational geometry (New Brunswick, NJ, 1989/1990), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 6, Amer. Math. Soc., Providence, RI, 1991, pp. 199–208,.

[L+04] J.A. De Loera, R. Hemmecke, J. Tauzer, and R. Yoshida, *Effective lattice point counting in rational convex polytopes*, see also http://www.math.ucdavis.edu/∼latte/, Journal of Symbolic Computation **38** (2004), 1273–1302.

[Pa94] C.H. Papadimitriou, *Computational Complexity*, Addison-Wesley, Reading, MA, 1994.

[St97] R.P. Stanley, *Enumerative Combinatorics. Vol 1*, Corrected reprint of the 1986 original. Cambridge Studies in Advanced Mathematics, vol. 49, Cambridge Univ. Press, Cambridge, 1997.

[V+04] S. Verdoolaege, R. Seghir, K. Beyls, V. Loechner, and M. Bruynooghe, *Analytical computation of Ehrhart polynomials: enabling more compiler analyses and optimizations*, see also http://www.kotnet.org/∼skimo/barvinok/, Proceedings of the 2004 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems (CASES 2004), 2004, pp. 248–258.

[Zi95] G. Ziegler, *Lectures on Polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995.