ON SAMPLING INTEGER POINTS IN POLYHEDRA

Igor Pak

Department of Mathematics,
Yale University,
New Haven, CT 06520
paki@math.yale.edu

April 8, 2000

Abstract. We investigate the problem of sampling integer points in rational polyhedra provided an oracle for counting these integer points. When dimension is bounded, this assumption is justified in view of a recent algorithm due to Barvinok [B1,B2,BP]. We show that in full generality the exactly uniform sampling is possible, when the oracle is called polynomial number of times. Further, when Barvinok's algorithm is used, poly-log number of calls suffices.

Introduction

Let $\mathrm{P} \subset \mathbb{R}^d$ be a rational polyhedra, where the dimension d is a fixed constant. Let $B = \mathrm{P} \cap \mathbb{Z}^d$ be the set of integer points in P. In a pioneer paper [B2], Barvinok presented an algorithm for computing $|B|$ in time polynomial in the size of the input. In sharp contrast with various approximation algorithms (see [DFK,KLS]), Barvinok's algorithm is algebraic, and by itself insufficient for sampling from B. In this paper we show how one can efficiently utilize advantages of this algorithm for uniform sampling from B.

The problem of uniform sampling of integer points in polyhedra is of interest in Computational Geometry as well as in Enumerative Combinatorics, Algebraic Geometry, and Applied Statistics (see [Br,DG,DKM,Sta,Stu]). There are numerous algorithms for uniform sampling of combinatorial objects (see e.g. [PW,W]), which often can be viewed as integer points in very special rational polyhedra. In statistics, one often need to obtain many independent uniform samples of the integer points in certain polyhedra (e. g. the set of contingency tables) to approximate a certain distribution on them (e. g. χ^2 distribution). We refer to [DE,DG] for references and details.

Let us note that Monte Carlo algorithms for *nearly uniform* sampling, based on Markov chain approach, have been of interest for some time. Remarkable polynomial time algorithms (including dimension!) have been discovered (see [DFK,KLS]).

Key words and phrases. Rational polyhedra, integer points, random sampling, Barvinok's algorithm.

Typeset by $\mathcal{AMS}$-TEX

These algorithms, however, work under certain "roundness" assumptions on polytopes and miss some "hard to reach" points. Theoretical results (see [DKM]) show hardness of uniform sampling in time polynomial in dimension. While the dimension of polytopes is often growing in cases of practical interest, it remains to be seen what can be done when the dimension is bounded.

By L everywhere below we denote the size of the input, and d will denote the dimension.

Theorem 1. *Let* $\mathrm{P} \subset \mathbb{R}^d$ *be a rational polytope, and let* $B = \mathrm{P} \cap \mathbb{Z}^d$. *Assume an oracle can compute* $|B|$ *for any* P *as above. Then there exists a polynomial time algorithm for sampling uniformly from* B *which calls an oracle* $O(d^2\, L^2)$ *times.*

Theorem 2. *In conditions of Theorem 1, there exists a polynomial time algorithm for sampling uniformly from* B *which calls Barvinok's algorithm* $O(d^2 \log L)$ *times.*

1. Uniform sampling

First we shall prove Theorem 1. Here is a general strategy. We will find a hyperplane H such that $\alpha = |B \cap H_+|/|B|, \beta = |B \cap H_-|/|B| \leq \frac{1}{2}$, where H_-, H_+ are halfspaces in $\mathbb{R}^d \setminus H$. Note that we can have $\gamma = |B \cap H|/|B| \geq \frac{1}{2}$, $\alpha+\beta+\gamma = 1$. Then sample a random variable with three outcomes, with probabilities α, β, γ. Depending on the outcome, reduce the problem to the smaller. Observe that either dimension drops, or the the number of integer points is reduced by a factor ≥ 2. Since the number of integer points is $\exp(d\,L)$, we need $O(d\,L)$ times to half it.

To find the hyperplane as above, consider all level hyperplanes $x_1 = C$, where x_i are coordinates in $V = \mathbb{R}^d$. Clearly, for some integer C this defines H as above. Now determine the constant C by binary search. Recall that C is bounded by $c_1 \leq C \leq c_2$, where c_1, c_2 are polynomial in $\exp(d\,L)$. Checking whether conditions $\alpha, \beta \leq \frac{1}{2}$ are satisfied requires two calls of an oracle for each constant to be tested, the total number of calls to half the polytope is $O(d\,L)$. Combining with the previous observation, this completes the proof of Theorem 1. $\square$

2. Using Barvinok's algorithm

The strategy is similar, but we will choose a desired constant C in a "smarter way", by utilizing the full power of Barvinok's algorithm.

Recall the idea of the algorithm in [B2] (see also [B1,BP]). Given a presentation of P by equations and inequalities, Barvinok computed $F(x) = F(x_1, \ldots, x_d; \mathrm{P})$ defined as

$$F(x_1, \ldots, x_d) = \sum_{m=(m_1,\ldots,m_d)\in B} x^m,$$

where $x^m = x_1^{m_1} \cdot \ldots \cdot x_d^{m_d}$. The solution is given in the form

$$(*) \qquad F(x) = \sum_{j \in J} \epsilon_j \, \frac{x^{a_j}}{(1 - x^b_{1,j}) \cdot \ldots \cdot (1 - x^{b_{d,j}})}\,,$$

where $\epsilon_j \in \{\pm 1\}$, $J = \{1, \dots, r\}$, and $a_j, b_{i,j} \in \mathbb{Z}^d$, are of size $L^{O(d)}$, polynomial in the size of the input. Now $|B| = F(1, \dots, 1)$, where the substitution is taken with care (cf. [DK]).

The real meaning of $(*)$ is that F is presented as a short alternating sum of the integer points of unimodular cones (with $det = \pm 1$). These cones originate in a_j and are mapped onto vertices of the polytope P. It is crucial that the number of cones $r = |J| = L^{O(d)}$.

Now we can present our algorithm which proves Theorem 2. For simplicity assume that $\mathrm{P} \in \mathbb{R}_+^d$, and has no facets parallel to $H = \{x_1 = 0\}$ (otherwise, one can always find a unimodular transformation of V to have that).

Let us orient all unimodular cones "upward", i.e. not to intersect H. Simply, for each $b_{i,j} \in H_-$ make a substitution $b'_{i,j} = -b'_{i,j}$, $\epsilon'_j = -\epsilon_j$, $a'_j = a_j - b_{i,j}$. Geometrically, this corresponds to flipping a cone in an appropriate cone with the same defining hyperplanes but different orientation. This is possible since the function $F(x) \equiv 0$ for sets containing lines (see part 4) of Theorem 3.1 in [BP]). Algebraically, this corresponds to substitution

$$\frac{1}{1 - z^{-1}} = \frac{-z}{1 - z}$$

for every $z = x^{b_{i,j}}$, $b_{i,j} \in H_-$.

Now observe that the volume $\mathrm{vol}(\mathrm{P} \cap \{x_1 \le C\})$ is piecewise polynomial in C, with the polynomial changing at first coordinate of vertices. Use binary search as in the previous section to determine between which of these the desired C lies (such that $\alpha, \beta \le \frac{1}{2}$ as in section 1.) The number of vertices is at most L^d, so $O(\log L)$ calls of an oracle suffices. One can simply pick random vertices, use oracle to determine the probabilities of restricting the polytope to either half, etc. With probability $\ge 1/2$ at most $3/4$ fraction of the points will remain in the half, so it will take $O(d \log L)$ iterations. At the end we obtain that the desired "random" point has to be sampled uniformly from a polytope $\mathrm{Q} = \mathrm{P} \cap \{c_1 \le x_1 \le c_2\}$.

Consider the structure of the polytope Q. Let $\mathrm{Q}_C = \mathrm{Q} \cap \{x_1 \le C\}$. From above, the volume $\mathrm{vol}(\mathrm{Q}_C)$ is polynomial in C degree C. Recall that we have presented all integer points in Q as an alternating sum of the integer points in the unimodular cones R_j, $j \in J$, since each cone R_j is chosen to have a compact intersection with a plane $\{x_1 = C\}$.

Fix one cone $\mathrm{R} = \{a + \mu_1 b_1 + \cdots + \mu_d b_d \,|\, \mu_i \in \mathbb{R}_+\}$, where $a, b_i \in \mathbb{Z}^d$. For simplicity, assume $a = 0$. Denote by M the sum of the first coordinates of b_i (all positive, from above). Observe that every integer point in R_{C-M} corresponds to a block of volume 1 in Q_c, which implies that

$$\left|\mathrm{R}_{C-M} \cap \mathbb{Z}^d\right| \le \mathrm{vol}(\mathrm{R}_C) \le \left|\mathrm{R}_{C+M} \cap \mathbb{Z}^d\right|.$$

By linearity, the above inequality holds for Q_C as well.

Now, the volume $\mathrm{vol}(\mathrm{R}_C)$ as a polynomial of degree d in C can be explicitly computed from a, b_i and c_1. Thus we obtain an explicit polynomial $f(C)$ for the volume of Q_C. Let $N = |\mathrm{Q} \cap \mathbb{Z}^d|$, and pick a random number $n \in \{1, \dots, N\}$. Estimate the unique solution C_0 of the equation $f(C) = n$ (up to the nearest integer). Then use binary search to determine the desired $C \in \{C_0 - M, \dots, C_0 + M\}$ (i.e. such that $\alpha, \beta \leq \frac{1}{2}$). This will require $O(\log L)$ oracle calls. Then proceed as in section 1.

Adding up the number of calls for Barvinok's algorithm, we conclude that for each of the d directions we need to call it $O(d \log L)$ times. This completes the proof of Theorem 2. □

3. Concluding remarks

It remains to be seen if Barvinok's algorithm is efficient in practice. In theory, it has $L^{O(d)}$ cost, which is perhaps excessive unless general assumptions are made. In particular, recall that one needs to calculate all vertices of the polyhedron when running Barvinok's algorithm. The main point of this note is to show that at a small additional cost one can use the algorithm to obtain uniform sampling as well.

Few simple observations to show that performance of our algorithm is somewhat better than we showed. First, recall that in section 2 all polytopes Q_C have the same combinatorial structure and thus covered by the second part of Theorem 4.4 in [BP]. Also, the estimate $O(d^2 \log L)$ is too conservative. One can make an argument that $O(d \log L)$ is enough when the hyperplane H is chosen appropriately. We do not believe additional analysis of our simple algorithm is justified.

Note that when faster approximation algorithms are available, one can use them in place of a counting oracle everywhere when determining which hyperplane to use. But the probabilities must be determined by the precise counting oracle since the errors will blow up otherwise.

Finally, when the function to be approximated on integer points is polynomial or exponential, one can use Barvinok's algorithm to obtain the exact result. In general case, however, our approach can be effective.

Acknowledgments

We would like to thank A. Barvinok for explaining his algorithm and helpful conversations. The author was partially supported by the NSF Postdoctoral Research Fellowship.

References

[B1] A. Barvinok, *Computing the volume, counting integer points, and exponential sums*, Discrete and Computational Geometry **10** (1993), 123–141.

[B2] 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.

[BP] A. Barvinok, J. Pommersheim, *An algorithmic theory of lattice points in polyhedra*, in New perspectives in algebraic combinatorics, Math. Sci. Res. Inst. Publ., 38, Cambridge Univ. Press, Cambridge, 1999, 91–147.

[Br] M. Brion, *Points entiers dans les polytopes convexes (Séminaire Bourbaki, Vol. 1993/94)*, Astérisque No. 227, (1995), Exp. No. 780, 4, 145–169.

[DE] P. Diaconis, B. Efron, *Testing for independence in a two-way table: new interpretations of the chi-square statistic. With discussions and with a reply by the authors.*, Ann. Statist. **13** (1985), 845–913.

[DG] P. Diaconis, A. Gangolli, *Rectangular arrays with fixed margins*, IMA series **72** (1995), Springer, 15–41.

[DFK] M. Dyer, A. Frieze, R. Kannan, *A random polynomial-time algorithm for approximating the volume of convex bodies*, J. ACM **38** (1991), 1–17.

[DK] M. Dyer, R. Kannan, *On Barvinok's algorithm for counting lattice points in fixed dimension*, Math. Oper. Res. **22** (1997), 545–549.

[DKM] M. Dyer, R. Kannan, J. Mount, *Sampling contingency tables*, Random Structures and Algorithms **10** (1997), 487–506.

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

[PW] J. Propp, D. B. Wilson, *How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph*, J. Algorithms **27** (1998), 170–217.

[Sc] A. Schrijer, *Theory of Integer and Linear Programming*, John Wiley, New York, NY, 1988.

[Sta] R. Stanley, *Combinatorics and Commutative Algebra*, Birkhouser, Boston, 1996.

[Stu] B. Sturmfels, *Equations defining toric varieties*, A.M.S. Proceeding of Symposia in Pure Mathematics **62** (1998), Providence, RI, 447–449.

[W] H. Wilf, *Combinatorial algorithms: an update,* CBMS-NSF Regional Conference Series in Applied Mathematics, 55., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1989.

[Z] G. Ziegler, *Lectures on Polytopes, Graduate Texts in Mathematics 152*, Springer, New York, 1995.