Efficient Random-Walk Methods for Approximating Polytope Volume

Ioannis Z. Emiris* Vissarion Fisikopoulos*

Abstract

We study the fundamental problem of computing the volume of a convex polytope given as an intersection of linear inequalities. We implement and experimentally evaluate practical algorithms for approximately approximating the polytope’s volume in high dimensions (one hundred). Our code is significantly faster than exact computation and more accurate, as well as faster, than existing approximation methods.

1 Introduction

A fundamental problem in discrete and computational geometry is to compute the volume of a convex body in general dimension or, more particularly, of a polytope. In the past 15 years, randomized algorithms for this problem have witnessed a remarkable progress. Starting with the breakthrough poly-time algorithm of [4], subsequent results brought down the exponent on the dimension from 27 to 4 [10]. However, the question of a practical implementation that handles general polytopes in high dimensions (a few hundred) had remained open. We tackle this question here by offering fast, accurate, public-domain software.

Convex bodies are typically given by a membership oracle, i.e., given point p, the oracle decides whether p lies in the body. A polytope P ⊆ Rd can also be represented as the convex hull of vertices (V-polytope) or, as is the case here, as the (bounded) intersection P := {x ∈ Rd | Ax ≤ b} of m halfspaces given by A ∈ Rm×d, b ∈ Rd (H-polytope); ∂P is its boundary.

Volume computation is #P hard for V- and for H-polytopes [5]. Several exact algorithms are surveyed in [2] and implemented in VINCI, the state-of-the-art software for exact volume computation, which, however, cannot handle general polytopes for dimension d > 15. An interesting challenge is the volume of the n-Birkhoff polytope, computed only for n ≤ 10 using highly specialized software (Sect. 3).

The landmark randomized poly-time algorithm in [4] approximates the volume of a convex body with high probability and arbitrarily small relative error. The best complexity, as a function of d, given a membership oracle, is O*(d) oracle calls [10], while a simpler, more geometric algorithm, which we shall use, requires O*(d5) calls [7]. O*(·) hides polylog factors in the argument. All approaches except [10] produce uniform point samples in successively larger convex subsets of P so as to approximate their volume.

Concerning software, in [9] they implement [10], focusing on variance-decreasing techniques, and an empirical estimation of mixing time. Very recently, a randomized algorithm in Matlab has been announced [1]; we show our software is faster and more accurate in sect. 3.

The key ingredient of all approaches is sampling points (almost) uniformly distributed in P. No simple sampling exists unless P is, e.g., a simplex, cube, or ellipsoid. Acceptance-rejection techniques are inefficient in high dimensions: the number of uniform points needed in a bounding box before finding one in P is exponential in d. A Markov chain is the only known method, and it may use (geometric) random walks such as the grid walk, the ball walk, and Hit-and-run [12]. The chain makes a large number of steps, before the generated point becomes distributed approximately uniformly. We focus on Hit-and-run which yields the fastest algorithms today.

Our contributions are multifold. Concerning point sampling we experimentally determine and set the length W of the random walk to O(d) which is much lower than the theoretical bounds and obtain a < 1% error in up to 100 dimensions (Sect. 3). Our emphasis is to exploit the underlying geometry. Our algorithm uses a sequence of co-centric balls, and samples points in their intersections with P (Sect. 2). Unlike previous methods, this forms a sequence of diminishing radii thus allowing us to only sample partial generations of points in each intersection with P, instead of sampling N points for each. Unlike most theoretical approaches, that use involved rounding procedures, we use a new simple method for iterative rounding that allows us to handle skinny polytopes efficiently (Sect. 2). Utilizing Coordinate Direction Hit-and-run, we design an oracle with O(m) amortized complexity (Sect. 2). We offer a series of experiments establishing that our code handles dimensions substantially larger than existing exact approaches.

http://www.cc.gatech.edu/~bcousins/Volume.html

---

*Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece. {emiris,vfisikop}@di.uoa.gr. This work is co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS - UOA (MIS 375891).
of 1% for \( d \leq 100 \), in about 20 min. This paper is an extended abstract of [6], which additionally offers a study that exploits duality to reduce the oracle to \( \varepsilon \)-nearest neighbor search.

2 The volume algorithm

Random walks, oracles, and sampling. Assume we possess procedure \texttt{Line}(p), which returns line \( \ell \) through point \( p \in P \subseteq \mathbb{R}^d \); \( \ell \) will be specified below. The main procedure of Hit-and-run is \texttt{Walk}(p, P, W), which reads in point \( p \in P \) and repeats \( W \) times: (i) run \texttt{Line}(p), (ii) move \( p \) to a random point uniformly distributed on \( P \cap \ell \). We shall consider two variants of Hit-and-run.

In Random Directions Hit-and-run (RDHR), \texttt{Line}(p) returns \( \ell \) defined by a random vector uniformly distributed on the unit sphere centered at \( p \). The vector coordinates are drawn from the standard normal distribution. RDHR generates a uniformly distributed point in \( 10^{30}O((d^2r)^2) \), or \( 10^{11}O((d^2r)^2) \) oracle calls starting at an arbitrary, or at a uniformly distributed point (aka warm start), respectively, where \( r \) is the ratio of the radius of the smallest enclosing ball over that of the largest enclosed ball in \( P \), and \( O(*) \) hides no constant [10].

In Coordinate Directions Hit-and-run (CDHR), \texttt{Line}(p) returns \( \ell \) defined by a random vector uniformly distributed on the set \( \{e_i: 0 \leq e_i = 0, \ldots, 1, 0, \ldots, 0\} \), \( i = 1, \ldots, d \). This is a continuous variant of the Grid walk. As far as the authors know, the mixing time has not been analyzed. In [6] we offer experimental evidence that CDHR is faster than RDHR and sufficiently accurate.

In contrast to other walks, Hit-and-run requires at every step the intersection of line \( \ell \) with \( \partial P \). In general, this reduces to binary search on the line, calling the membership oracle at every step. For H-polytopes, the intersection is obtained by a boundary oracle; for this, we employ ray-shooting with respect to the \( m \) facet hyperplanes. We focus on CDHR, therefore we may suppose \( \ell \) is vertical. Let us consider \texttt{Walk}(p_0, P, W) and vertical line \( \ell = \{x \in \mathbb{R}^d : x = \lambda e + p_0\} \), where \( p_0 \in \mathbb{R}^d \) lies on \( \ell \), and \( \lambda \) is the vertical direction. We compute the intersection of \( \ell \) with the \( i \)-th hyperplane \( ax = b_i, a_i \in \mathbb{R}^d, b_i \in \mathbb{R} \), namely \( p_i := p_0 + \frac{b_i - a_i e}{a_i} v, i = 1, \ldots, m \). We seek points \( p^+, p^- \) at which \( \ell \) intersects \( \partial P \), namely \( p^+ v = \min_{1 \leq i \leq m} \{b_i v | p_i v \geq 0\} \), \( p^- v = \max_{1 \leq i \leq m} \{b_i v | p_i v \leq 0\} \). This is computed in \( O(md) \) arithmetic operations. In practice, only the \( \lambda^k \) are computed, where \( p^k = p_0 + \lambda^k v \). However, in CDHR, after the computation of the first pair \( p^+, p^- \), all other pairs can be computed in \( O(m) \) arithmetic operations. This is because two sequential points produced by the walk differ only in one coordinate.

Given polytope \( P \subseteq \mathbb{R}^d \) and approximation factor \( \epsilon > 0 \), the volume algorithm executes sandwiching and Multiphase Monte Carlo (MMC) [11].

Rounding and sandwiching. There is an abundance of methods in literature for sandwiching (cf. [11] and references therein). The goal is to compute ball \( B \) and scalar \( \rho > 1 \) such that \( B \subseteq P \subseteq \rho B \). However, here we develop a simpler method that instead of computing \( \rho \) such that \( \rho B \) covers \( P \), computes \( B' \) such that \( B' \cap P \) contains almost all the volume of \( P \). Our method handles efficiently skinny polytopes where \( \rho \) is large (Sect. 3).

To this end we perform rounding of \( P \). We sample a set \( S \) of \( O(n) \) random points in \( P \) using the random walk methods described above. Then we approximate the minimum volume ellipsoid \( E \) that covers \( S \), and satisfies the inclusions \( \frac{1}{1 + \log d + \log n} E \subseteq \text{conv}(S) \subseteq \text{E} \), in time \( O(n(d^2e^{-1} + \log d + \log n)) \) [8]. Let us write

\[
E = \{x \in \mathbb{R}^d | (x - c)'^T E (x - c) \leq 1\} \quad (1)
\]

where \( E \subseteq \mathbb{R}^{d \times d} \) is a positive semi-definite matrix and \( L^T L \) its Cholesky decomposition. By substituting \( x = (L^T)^{-1} y + c \), \( E \) we map the ellipsoid to the ball \( \{y \in \mathbb{R}^d | y^T y \leq 1\} \). Applying this transformation to \( P \) we have \( P' = \{y \in \mathbb{R}^d | A(L^T)^{-1} y \leq b - Ac \} \) which is the rounded polytope, where \( \text{vol}(P) = \text{det}(L^T)^{-1} \text{vol}(P') \). We iterate this procedure until the variance of the set of ellipsoid axes reaches some user-defined threshold.

For sandwiching we first compute the Chebychev ball \( B(c, r) \) of \( P \), i.e. the largest inscribed ball in \( P \). It suffices to solve the LP: \{maximize \( R \), subject to: \( A_i x + R \|A_i\|_2 \leq b_i, i = 1, \ldots, m, R \geq 0\)\}, where \( A_i \) is the \( i \)-th row of \( A \), and the optimal values of \( R \) and \( x \in \mathbb{R}^d \) yield, respectively, the radius \( r \) and the center \( c \) of the Chebychev ball.

Then we may compute a uniform random point in \( B(c, r) \) and use it as a start to perform a random walk in \( P \), eventually generating \( N \) random points. Now, set \( \rho \) to be the largest distance between each of the \( N \) points and \( c \); this defines a (approximate) bounding ball. Finally, define the sequence of balls \( B(c, 2^\alpha i), i = \alpha, \alpha + 1, \ldots, \beta \), where \( \alpha = \lfloor d \log r \rfloor \) and \( \beta = \lceil d \log \rho \rceil \).

Multiphase Monte Carlo (MMC). MMC constructs a sequence of bodies \( P_i := P \cap B(c, 2^\alpha i), i = \alpha, \alpha + 1, \ldots, \beta \), where \( P_0 = B(c, 2^\alpha i) \subseteq B(c, r) \) and \( P_\beta \) almost contains \( P \). Then it approximates \( \text{vol}(P) \) by the telescopic product \( \text{vol}(P_i) \prod_{i=0}^{\beta} \frac{\text{vol}(P_i)}{\text{vol}(P_{i+1})} \), where \( \text{vol}(P_0) = 2\pi^{d/2}(2\log r)^d/d! \).

This reduces to estimating the ratios \( \text{vol}(P_i)/\text{vol}(P_{i+1}) \), which is achieved by generating \( N \) uniformly distributed points in \( P_i \) and by counting how many of them fall in \( P_{i+1} \).

For point generation we use random walks as in Sect. 2. We set the walk length \( W = \lfloor 10 + d/10 \rfloor \).
partial generations of random points for every new
Complexity. It reduces it by a constant raised to
reduces to testing whether
Assuming algorithms using a sequence of concentric balls, like [7].
lie in

\( D \)

\( P \)

compute vol(\( \cdot \))

\( i > \alpha \)

\( \log(\rho) \)

\( \rho \)

\( \Delta-d \)

\( \Delta-d \)-d: product of two simplices, i.e \((p,p') \in \mathbb{R}^{2d} | p \in \Delta-d, p' \in \Delta-d \),

\( s\)-cube-d: \( \{ x = (x_1, \ldots, x_d) | x_1 \leq 100, x_1 \geq -100, x_i \geq -1, x_i \in \mathbb{R} i = 2, \ldots, d \} \), rotated by \( 30^\circ \) in the plane defined by the first two coordinate axes,

\( \mathcal{B}(n) \): the \( n \)-Birkhoff polytope (defined below).

Each experiment is repeated 100 times. We keep track of the \textit{min} and the \textit{max} computed values, the mean \( \mu \), the standard deviation and the mean error of approximation \((\text{vol}(P) - \mu)/\text{vol}(P)\). Our method is more accurate than indicated by the theoretical bounds in [7]. In particular, in all experiments all computed values are contained in the interval \((1 - \epsilon)\text{vol}(P), (1 + \epsilon)\text{vol}(P)\), while theory guarantees only \( 75\% \) of them. In general our experimental results show that our software can approximate the volume of general polytopes up to dimension 100 in less than 2 hours with mean approximation error at most 2\% (cf. Table 1).

We set \( W = [10 + d/10] \). Our experiments indicate that, with this choice, \((\text{vol}(P) - \mu)/\text{vol}(P)\) is \( < 2\% \) up to \( d = 100 \) (Table 1). Moreover, for higher \( W \) the improvement in accuracy is not significant, which supports the claim that asymptotic bounds are unrealistically high.

\[ \frac{2}{\\text{VolEsti}} \]

\[ \rho \text{-cube-} \]

\[ \text{vol}(P) \]

\[ \mathcal{B}(n) \]

\[ \mathbb{R}^{2d} \]

\[ \Delta-d \]

\[ p \in \Delta-d, p' \in \Delta-d \]

\[ s\text{-cube-} \]

\[ x = (x_1, \ldots, x_d) \]

\[ x_i \geq -1, x_i \in \mathbb{R} i = 2, \ldots, d \]

\[ 30^\circ \]

\[ \Delta-d \]

\[ \Delta-d \]-d: product of two simplices, i.e \((p,p') \in \mathbb{R}^{2d} | p \in \Delta-d, p' \in \Delta-d \),

\[ s\text{-cube-} \]

\[ x = (x_1, \ldots, x_d) \]

\[ x_i \geq -1, x_i \in \mathbb{R} i = 2, \ldots, d \]

\[ 30^\circ \]

\[ \Delta-d \]

\[ \Delta-d \]-d: product of two simplices, i.e \((p,p') \in \mathbb{R}^{2d} | p \in \Delta-d, p' \in \Delta-d \),

\[ s\text{-cube-} \]

\[ x = (x_1, \ldots, x_d) \]

\[ x_i \geq -1, x_i \in \mathbb{R} i = 2, \ldots, d \]

\[ 30^\circ \]

\[ \Delta-d \]

\[ \Delta-d \]-d: product of two simplices, i.e \((p,p') \in \mathbb{R}^{2d} | p \in \Delta-d, p' \in \Delta-d \),

\[ s\text{-cube-} \]

\[ x = (x_1, \ldots, x_d) \]

\[ x_i \geq -1, x_i \in \mathbb{R} i = 2, \ldots, d \]

\[ 30^\circ \]
To experimentally test the effect of rounding we construct skinny hypercubes s-cube-d. We rotate them to avoid CDHR taking unfair advantage of the degenerate situation where the long edge is parallel to an axis. Table 1 shows that rounding reduces approximation error by 2 orders of magnitude.

We test against VINCI 1.0.5 [2] (Table 1). For all inputs, there is a threshold dimension for which VINCI takes too much time (e.g. from 4 to 100 times smaller error on cube-20 and 11% for vol(B) 12). We decrease $\epsilon$ from 0.5 [2] (Table 1). For all inputs, there is a threshold dimension for which VINCI takes too much time (e.g. from 4 to 100 times smaller error on cube-20 and 11% for vol(B) 12).

Testing the most relevant approximation method implemented in Matlab (cf. Sect. 1) with default options and $\epsilon = 0.1$, our implementation runs at least 2 times faster and returns significantly more accurate results, e.g. from 4 to 100 times smaller error on cube-d when $d > 70$, and from 5 to 80 times on Birkhoff polytopes (Table 1).

For example, for $B_{10}$ Matlab and VolEsti compute $\frac{\text{vol}(B_{10})}{\text{vol}(P)} = 0.292$ and 0.293 respectively in 3437.74 and 3600.20 secs respectively and for cube-100 compute $\frac{\text{vol}(B_{100})}{\text{vol}(P)} = 0.0357$ and 0.0081 in 3805.37 and 1285.08 secs respectively.

The $n$-th Birkhoff polytope $B_n = \{(x_i) \in \mathbb{R}^{n \times n} \mid x_{ij} \geq 0, \sum x_{ij} = 1, \sum x_{i1} = n\}$, is also described as the polytope of the perfect matchings of the complete bipartite graph $K_{n,n}$. In [1], they present a complex-analytic method for this volume, implemented in package birkhoff, which has managed to compute $\text{vol}(B_0)$ in parallel execution, which corresponds to a single processor running at 1 GHz for almost 17 years. Our software computes the volume of polytopes up to $B_{10}$ in < 1 hour with mean error of ≤ 2% (Table 1). We decrease $\epsilon$ and obtain an error of 0.7% for vol(B3), in 6 hours, i.e., with two correct digits. More interestingly, using $\epsilon = 0.5$ we compute an approximation as well as an interval for $\text{vol}(B_{11})$, $\text{vol}(B_{12})$, whose exact values are unknown (Table 1).

**References**


