Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 4 (2011) 1403–1411

Procedia Computer Science

International Conference on Computational Science, ICCS 2011

A Monte Carlo Method for High-Dimensional Volume Estimation and Application to Polytopes

Uwe Jaekel*

Department of Mathematics and Technology, University of Applied Sciences Koblenz, 53424 Remagen, Germany

Abstract

A direct Monte Carlo method for volume estimation of star-shaped or convex domains is presented, and is generalized to a Markov Chain Monte Carlo method for high-dimensional problems. The direct approach itself, which is closely related to a method proposed by Fok and Crevier, is already applicable to some moderately high-dimensional problems. The combination with a Markov Chain Monte Carlo method and nested sampling or thermodynamic integration extends its scope considerably. Applications to the volume estimation for high-dimensional polytopes are presented, and the method is tested using exact results for the first 10 Birkhoff polytopes.

Keywords: Volume estimation, Monte Carlo methods, Thermodynamic integration, Polytopes, Birkhoff polytopes

1. Introduction

Estimating the volume of a high-dimensional domain is a difficult problem with important applications in many fields. The first random polynomial time algorithm for approximating the volume of a convex body has been proposed by Dyer, Frieze and Kannan [1] with a very high power of the dimension d, that could be reduced in later works. Currently, the best available algorithm [2] is based on a multiphase Markov chain Monte Carlo approach that requires $O^*(d^4)$ calls of a membership oracle, where $O^*(d^4)$ means $O(d^4)$ up to polynomial factors in $\log(d)$.

In this paper, an alternative Monte Carlo algorithm is constructed from a representation of the volume of a convex or star shaped domain as an integral over the unit sphere. This allows to construct a direct Monte Carlo method that already leads to good results for some interesting test problems with fewer than about 20 dimensions. For higher-dimensional problems, however, the curse of dimensionality is reflected in the occurrence of very sharp peaks of the integrand, such that it becomes necessary to combine this method with importance sampling techniques.

To overcome the limitations of the direct Monte Carlo approach, a generalization of this method to a Markov Chain Monte Carlo (MCMC) approach is presented that is based on the interpretation of the integrand as a probability distribution with an unknown normalization coefficient. The Metropolis-Hastings method allows to sample preferably close to peaks of the integrand without knowledge of the normalization factor. In order to find the latter, the Metropolis-Hastings method is combined with a thermodynamic integration method. Thermodynamic integration is a

* *Email address:* jaekel@rheinahrcampus.de (Uwe Jaekel)

Selection and/or peer-review under responsibility of Prof. Mitsuhisa Sato and Prof. Satoshi Matsuoka
doi:10.1016/j.procs.2011.04.151

method originating in statistical physics and can favorably be combined with parallel tempering approaches that allow to overcome problems arising from the possibly highly multi-modal density over the unit sphere.

An important application of volume estimation, relevant e.g. in theoretical and systems biology ([4], [5], [6]) is volume estimation for high-dimensional polytopes. The methods developed in this paper were tested with the first 10 Birkhoff polytopes, for which exact results are known ([7]-[9]). The Birkhoff polytopes are sets of doubly stochastic matrices occurring in a wide range of applications from combinatorial optimization to quantum information.

2. A direct Monte Carlo method for volume estimation

2.1. The general idea

A direct Monte Carlo method for volume estimation in high-dimensional spaces has been proposed by Fok and Crevier [3], assuming that it is possible to find for every direction $\mathbf{e}$ the point at which a straight line starting from an interior point of the domain with direction $\mathbf{e}$ hits the boundary. In many important applications this can be done efficiently, either in closed form or by iterative methods.

We derive a related, but not completely equivalent method using Gauss' divergence theorem, assuming that $\Omega \subset \mathbb{R}^d$ is a compact domain with piecewise smooth boundary. Let $\mathbf{F}$ be a vector field with constant divergence $\mathrm{div}\mathbf{F} = 1$. A natural – but not necessarily optimal – choice is $\mathbf{F}(\mathbf{x}) = \frac{1}{d}\mathbf{x}$. Then the volume V of the domain can be computed using the divergence theorem

$$V = \int_\Omega \mathrm{d}V = \int_K \mathrm{div}\mathbf{F}\,\mathrm{d}V = \int_{\partial\Omega} \mathrm{div}\mathbf{F}\cdot\mathbf{n}\,\mathrm{d}\sigma, \tag{1}$$

where $\partial\Omega$ is the boundary of the domain, $\mathbf{n}(\mathbf{x})$ is the outer normal vector at the point $\mathbf{x}$ of the boundary, $\mathrm{d}\sigma$ is the $(d-1)$-dimensional volume element on the boundary, and $\mathbf{a}\cdot\mathbf{b}$ denotes the scalar product of the two vectors $\mathbf{a}$ and $\mathbf{b}$. For the moment, we assume that K is star shaped around an interior point $\mathbf{x_0}$ located inside Ω. Then the volume can be estimated as a Monte Carlo integral over $\mathbf{F}\cdot\mathbf{n}$ using the following algorithm:

1. Draw K vectors $\mathbf{e}_k, k = 1\ldots K$ uniformly distributed from the unit sphere.
2. Starting from $\mathbf{x_0}$, move along direction $\mathbf{e}_k$ until the boundary is hit, i.e. compute the smallest parameter $t_k > 0$ such that $\mathbf{x}_k = \mathbf{x_0} + t\mathbf{e}_k$ lies on the boundary.
3. Compute the normal vector $\mathbf{n}_k$ and the value $\mathbf{F}_k$ of the vector field $\mathbf{F}$ at $\mathbf{x}_k$.

The volume element on the unit sphere is scaled by the projection to the boundary at $\mathbf{x}_k$ by a factor of $t^{d-1}/\mathbf{n}_k\cdot\mathbf{e}_k$. The surface hypervolume of the n-dimensional unit sphere is $A_d = 2\Gamma(d/2)/\pi^{d/2}$ and we find for the Monte Carlo estimator

$$\hat{V}_K = \frac{2\Gamma(d/2)}{N\pi^{d/2}} \sum_{k=1}^{K} \frac{\mathbf{F}_k\cdot\mathbf{n}_k}{\mathbf{e}_k\cdot\mathbf{n}_k} t_k^{d-1} \to V \quad \text{for } K\to\infty. \tag{2}$$

2.2. Choice of the vector field and the Fok-Crevier estimator

For the choice $\mathbf{F}(\mathbf{x}) = \frac{1}{d}\mathbf{x}$, the estimator becomes

$$\hat{V}_K = \frac{2\Gamma(d/2)}{dK\pi^{d/2}} \sum_{k=1}^{K} \frac{\mathbf{x}_k\cdot\mathbf{n}_k}{\mathbf{e}_k\cdot\mathbf{n}_k} t_k^{d-1}. \tag{3}$$

The origin of the co-ordinate system can be located anywhere. However, in order to minimize the variance and to avoid roundoff errors, the projection $\mathbf{x}_k\cdot\mathbf{n}_k$ should not change its sign, i.e. the origin should lie inside Ω. If the origin is shifted to the point $\mathbf{x_0}$, the search direction vector $\mathbf{e}_k$ is parallel to the boundary vector $\mathbf{x}_k = t_k\mathbf{e}_k$, such that the estimator reduces to

$$\hat{V}_K = \frac{2\Gamma(d/2)}{dK\pi^{d/2}} \sum_{k=1}^{K} \frac{t_k\mathbf{e}_k\cdot\mathbf{n}_k}{\mathbf{e}_k\cdot\mathbf{n}_k} t_k^{d-1} = \frac{2\Gamma(d/2)}{dK\pi^{d/2}} \sum_{k=1}^{K} t_k^{d}. \tag{4}$$

This is the estimator proposed by Fok and Crevier [3], which can therefore be regarded as a special case of the estimator in Eq. (2). Note, however, that the estimator in Eq. (4) can be derived without assumptions on the smoothness of the boundary $\partial\Omega$.

It remains to be shown that there are reasonable choices for the vector field $\mathbf{F}$ and the location of the reference point $\mathbf{x_0}$ other than those leading to the estimator by Fok and Crevier. In order to see that, consider a thin rectangle $I = \{(x, y)|0 \le x \le 1 \wedge 0 \le L\}$ where $L \gg 1$, and assume that the integral shall be computed using an importance sampling method (such as the Markov Chain Monte Carlo method developed in section 3), in a variant where $\mathbf{x_0} = (0, 0)$, and the integral is only over the intersection of the unit sphere with the first quadrant. For the choice leading to the Fok-Crevier estimator, the MCMC method will sample values for t ranging from 1 to $\sqrt{L^2 + 1}$. If we choose $\mathbf{F}(x, y) = (0, y)$, the value of $\mathbf{F} \cdot \mathbf{n}$ will only be non-zero on the upper boundary $\{(x, y)|0 \le x \le 1 \wedge y = L\}$. Therefore, importance sampling will avoid the right boundary completely, and sample only values for t ranging from L to $\sqrt{L^2 + 1}$, leading to a much smaller variance of the estimator.

2.3. Generalizations

2.3.1. Membership oracles and convex domains

Many of the available methods for volume estimation only require the existence of a membership oracle, which for the general case is a much less restrictive requirement than an algorithm that computes all intersections of a straight line with the boundary of the inspected domain. However, for convex domains, starting from an interior point $\mathbf{x_0}$, a membership oracle can be used to find the intersection points efficiently by line search methods. A membership oracle is therefore not a less restrictive requirement for convex domains.

2.3.2. Integrations over domains that are not star shaped

For domains that are not convex or star shaped, the generalization is straightforward in principle, and analogous to the construction in [3], provided there is an algorithm that yields all intersections of a straight line with the boundary of the domain as well as the oriented normal vectors in these boundary points. This requirement is obviously fulfilled for polytopes. However, the non-negativity of the integrand will be lost, which makes the application of the MCMC generalization developed in section 3 more difficult.

2.4. Application to convex polytopes

2.4.1. Polytopes without equality constraints

A convex polytope can be characterized as the solution set of multiple inequalities:

$$P = \left\{ \mathbf{x} \in \mathbb{R}^d \middle| \mathbf{Ax} \le \mathbf{b} \right\}$$

where $\mathbf{A}$ is an $r \times d$-matrix, $\mathbf{b} \in \mathbb{R}^r$, and the inequality is to be understood componentwise. In this case, the value t where $\mathbf{x} = \mathbf{x_0} + t\mathbf{e}$ hits the boundary is the minimal value $t > 0$ above which one of the constraints is violated. If $\mathbf{a}_j$ is the j-th row vector of $\mathbf{A}$, and b_j the j-th component of $\mathbf{b}$, this value can be computed from

$$t^{(j)} = \frac{b_j - \mathbf{a}_j \cdot \mathbf{x_0}}{\mathbf{a}_j \cdot \mathbf{e}}, \qquad j = 1, \ldots r$$

$$t = \min\left(\left\{ t^{(1)}, \ldots, t^{(r)} \right\} \cap \left\{ t^{(j)} \middle| t^{(j)} > 0 \right\} \right).$$

The normal vector of the surface that has been hit is the corresponding row vector of $\mathbf{A}$.

2.5. Polytopes with equality constraints

Many polytopes useful in practical applications are defined via mixed inequality and equality constraints:

$$P = \left\{ \mathbf{x} \in \mathbb{R}^d \middle| \mathbf{A} \le \mathbf{b} \wedge \mathbf{Cx} = \mathbf{d} \right\}$$

where $\mathbf{A}$ is an $s \times d$-matrix, $\mathbf{d} \in \mathbb{R}^s$. Interesting special cases are transportation or flux polytopes that play an important role, e.g. for biomedical applications in the description of metabolic networks [5], [4]. The algorithm can be adapted

easily to this problem by the computation constructing an inner point of the polytope, and then restricting the search directions to the null space of C.

Famous representatives of this class are the Birkhoff polytopes which serve as test problems in this paper. The n-th Birkhoff polytope $\mathcal{B}_n$is formed by the set of doubly stochastic $n \times n$ matrices, i.e. matrices with non-negative entries where all row sums and all column sums are equal to 1. The Birkhoff polytopes therefore form an $(n-1)^2$-dimensional subspace of the space of $n \times n$ matrices. $\mathcal{B}_n$ has n^2 facets and, as it can be characterized as the convex hull of the n-dimensional permutation matrices, $n!$ vertices. The Birkhoff polytopes were selected as non-trivial reference cases since exact results have been derived for the volumes of the first 10 Birkhoff polytopes [7],[8], [9] in the case of $\mathcal{B}_{10}$ by a long and massive computation on a workstation farm. The volume of $\mathcal{B}_{10}$ is

$$\mathrm{vol}\mathcal{B}_{10} \approx 8.8 \cdot 10^{-46}$$

relative to the 81-dimensional unit hypercube. The naive Monte Carlo approach by sampling from the unit hypercube and counting points hitting the polytope is obviously hopeless in this case.

2.6. Application of the direct approach to high-dimensional polytopes

The direct algorithm can be implemented in a few lines of MATLAB code. There are only two implementation details one should care about:

1. For this straightforward implementation, the reference point $\mathbf{x_0}$ obviously should not be located close to a vertex: If the vertex is, for instance, the intersection of d orthogonal hyperplanes, the probability that a direction vector sampled from a uniform distribution on the unit sphere does not point to one of the nearby facets converges to 2^{-d}. With a probability of of $1-2^{-d}$, the distance t to the boundary will be tiny and give a negligible contribution to the integral, as the integrand is proportional to t^d. If available, an estimation for the center of the polytope is a good choice. Alternatively, since the expectation value does not depend on the location of $\mathbf{x_0}$, it is possible to let the reference point walk through the polytope. Good results can be achieved by keeping track of the maximum value of the integrand on the sample, and moving the reference point to $\mathbf{x_0} + \frac{1}{2}t_k\mathbf{e}_k$ whenever the integrand exceeds a few percent of the maximum.
2. It should be controlled if the polytope is thin in one or more directions. This can be achieved by keeping track of the covariance matrix C of the sampled boundary points. Small eigenvalues reveal "thin directions" in form of the corresponding eigenvectors. These can be compensated by rescaling the coordinates using the eigenvectors as a new basis.

The direct method described above was tested with standard simplices and Birkhoff the polytopes $\mathcal{B}_1, \ldots, \mathcal{B}_{10}$. The volume of the d-dimensional standard simplex

$$\Delta_{d-1} = \left\{ \mathbf{x} \in \mathbb{R}^d \middle| 0 \le x_i \le 1 \forall i = 1 \ldots d \wedge \sum_{i=1}^{d} x_i = 1 \right\}$$

is $1/d!$. For $d = 30$, this is about $3.77 \cdot 10^{-33}$, so again the naive Monte Carlo approach counting hits from uniformly distributed samples in the unit hypercube is not applicable.

In many applications, such as the estimation of the evidence of a model or the evidence against a hypothesis, it is sufficient to estimate the order of magnitude of the results. If we regard the method (somewhat arbitrarily) as applicable if it gives the correct order of magnitude, i.e. a logarithmic ($\log_{10}$) difference between the exact result and the estimator less than $1/2$ and a corresponding confidence interval, within a computation time of about 1 hour on a single core of a 2GHz PC, the algorithm seems to work up to $d \approx 35 \ldots 45$ for the simplex, with much better accuracy of course for lower dimensions. The volume of a 10-dimensional standard simplex ($\mathrm{vol}\Delta_9 \approx 2.76 \cdot 10^{-7}$), for instance, can be computed with two correct decimal places within a few minutes. These results were consistent with the results for the Birkhoff polytope: The order of magnitude is correct for $\mathcal{B}_6$ (25 dimensions) after a few minutes, while the algorithm needs hours for $\mathcal{B}_7$ (36 dimensions). These results can be improved of course by variance reduction techniques. Nevertheless it is worthwhile to note that already this simple method gives good results for moderately high dimensions, at least in applications that don't require too many computations on varying domains. In general, however, for more than 20 dimensions, one would certainly wish for lower variance estimators, which we shall focus on in the following section.

2.7. Importance sampling

For polytopes, the weight t^d of the integrand is obviously sharply peaked at the vertices. Therefore, importance sampling methods can be tailored to the polytope that sample preferably around the vertices of the polytope. For the Birkhoff polytope this is particularly easy since its vertices are many, but known, and all have the same distance to the matrix $a_{ij} = 1/n \quad \forall 1 \le i, j \le n$. However, in order to generalize the method also for higher-dimensional domains that are not bounded by hyperplanes (such as level sets of a convex function), we will focus on a Markov Chain Monte Carlo variant.

3. A generalization based on Markov Chain Monte Carlo methods

If the domain does not have an almost circular shape, the values of the integrand, that contains a factor proportional to t^d will fluctuate over many orders of magnitude for high dimensions, and will be sharply peaked around its maxima. If the vector field is chosen such that $\mathbf{F} \cdot \mathbf{x}$ is non-negative, as it is the case e.g. in the Fok-Crevier estimator in Eq. (4), we can interpret the integrand normalized to the volume

$$p(\mathbf{x}) := \frac{1}{V} \frac{\mathbf{F} \cdot \mathbf{n}}{\mathbf{e} \cdot \mathbf{n}} t_k^{d-1} \tag{5}$$

as a probability density. Expectations under this density can be estimated using Markov Chain Monte Carlo methods such as the Metropolis-Hastings algorithm in which the density has to be known only up to a constant factor. The problem now is that the normalization constant is the volume that we wish to compute, and therefore remains unknown. Several approaches to the problem of estimating normalization constants, such as nested sampling or bridging or annealing techniques, have been suggested in different disciplines ([11], [12]). For the purpose of this paper, we choose a method called thermodynamic integration due to its interpretation in statistical physics. The method is equivalent to bridge sampling.

3.1. Thermodynamic integration

In order to express the volume as an expectation under the density in Eq. (5), we introduce a family of densities

$$p_\beta(\mathbf{x}) := \frac{1}{Z(\beta)} f(\mathbf{x})^\beta = \frac{1}{Z(\beta)} \exp\left(\beta \ln f(\mathbf{x})\right), \tag{6}$$

where

$$Z(\beta) = \int_{S^{d-1}} \exp\left(\beta \ln f(\mathbf{x}(\omega))\right) \, \mathrm{d}\omega, \tag{7}$$

where S^{d-1} is the d-dimensional unit sphere. In statistical physics, β is interpreted as an inverse temperature. Obviously, for $\beta = 0$, the integrand is 1, and $Z(0)$ is the surface hypervolume of S^{d-1}, such that

$$\begin{aligned} Z(0) &= A_d = \frac{2\Gamma(d/2)}{\pi^{d/2}} \\ Z(1) &= V \end{aligned} \tag{8}$$

On the other hand,

$$\begin{aligned} \ln \frac{Z(1)}{Z(0)} &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\beta} \ln Z(\beta) \, \mathrm{d}\beta \\ &= \int_0^1 \frac{1}{Z(\beta)} \int_{S^{d-1}} \ln f(\mathbf{x}(\omega)) \exp\left(\beta \ln f(\mathbf{x})\right) \, \mathrm{d}\beta \\ &= \int_0^1 \int_{S^{d-1}} \ln f(\mathbf{x}(\omega)) p_\beta(\mathbf{x}(\omega)) \, \mathrm{d}\omega \, \mathrm{d}\beta \\ &= \int_0^1 E_{p_\beta}\left[\ln f\right] \, \mathrm{d}\beta \end{aligned} \tag{9}$$

where $E_{p_\beta}[\ln f]$ is the expectation value of $\ln f$ under the density p_β. The integrand can be estimated using the Metropolis-Hastings algorithm for any given value of β, such that the integral can be computed numerically. A simple choice for the proposal distribution is a random walk step on the unit sphere, such that for any given β, a simple algorithm for the evaluation of the integrand in a fixed number K of steps starting with a search direction $\mathbf{e}_0$ with a value $f_0 = f(\mathbf{x}_0)$ of the integrand is as follows:

1. **for** $k = 1 \ldots K$
2. Draw a direction vector $\mathbf{e}'$ uniformly distributed on the unit sphere.
3. Set $\mathbf{e}'' = (\mathbf{e}_{k-1} + \delta\xi\mathbf{e}')$ where $0 < \delta < 1$ is a fixed constant, and $\xi \in U(0,1)$ is a uniformly distributed random num
4. Set $\mathbf{e}''' = \mathbf{e}''/\|e''\|$.
5. Set f' to the value of f in direction $\mathbf{e}'''$.
6. Accept e''' as new search direction $\mathbf{e}_k$ with probability $\min(1, f'^\beta/f_{k-1}^\beta)$, otherwise let the search direction remain unchanged, i.e. $\mathbf{e}_k = \mathbf{e}_{k-1}$.
7. Set f_k to the value of f in direction $\mathbf{e}_k$.
8. **end for.**
9. Let $\hat{I}_K(\beta) = \frac{1}{K}\sum_{k=1}^K f_k^\beta$.

As usual, one should perform a bigger number of Metropolis steps (2-6) before $\hat{I}_K$ is sampled, in order to make sure that the starting direction $\mathbf{e}_0$ is not too far away from a peak. The volume can then be estimated from

$$V = Z(1) = A_d \exp\left(\int_0^1 E_{p_\beta}[\ln f]\ \mathrm{d}\beta\right), \tag{10}$$

and the MCMC estimator for V is

$$\hat{V}_K^{\mathrm{MCMC}} = A_d \exp\left(\int_0^1 \hat{I}_K(\beta)\,\mathrm{d}\beta\right) \tag{11}$$

where the integral $\int_0^1 \hat{I}_K(\beta)\,\mathrm{d}\beta$ is evaluated numerically.

For a general high-dimensional polytope, the distribution will be highly multimodal, and so sharply peaked that transitions of the Markov chains from one peak to another are very unlikely. Therefore, if one wants to use the simple algorithm above, it is necessary to restart the algorithm with different random start directions $\mathbf{e}_0$ repeatedly, and average over the resulting values of $\hat{I}_K$. A much better alternative is to evaluate the integral using path sampling techniques that elegantly combine transitions between the peaks at lower values of β and exploration of the details of f at $\beta \approx 1$ ([11], [12]).

3.2. Parallel tempering

Parallel tempering [13] is a technique to keep Markov chains from getting caught in local modes. Not necessarily related to thermodynamic integration, the idea is to run Markov chains at different "inverse temperatures" β, with densities defined as in the previous section, and allow the exchange states between different temperatures $0 = \beta_1 < \beta_2 < \ldots < \beta_M = 1$. The MCMC method described above is augmented by additional "replica exchange" steps that accept a proposed swap of the state $\mathbf{e}^{(j)}$ at inverse temperature β_j and the state $\mathbf{e}^{(i)}$ at another β_i (usually $i = j-1$) with probability

$$\min\left(1, \exp\left((\beta_j - \beta_i)\left(H(\mathbf{e}^{(i)}) - H(\mathbf{e}^{(j)})\right)\right)\right)$$

where $H = \ln f$. This rule guarantees detailed balance and therefore under mild restrictions convergence of the Markov chains at all temperatures. Markov chains that would otherwise be trapped for large values of β close to a local maximum of the density f, are released by parallel tempering, since transitions between local modes are very likely for sufficiently small values of β. Often, one is interested only in the distribution for $\beta = 1$, but for the case of thermodynamic integration, all copies at all temperatures are contributing to the evaluation of the integral.

3.3. Numerical results

Table 1 shows the results for a straightforward MATLAB implementation of the parallel tempering MCMC method described above, applied to the Birkhoff polytopes $\mathcal{B}_5 \ldots \mathcal{B}_{10}$, on a single core of a 2.67 GHz workstation. 100 Copies of a Markov chain at different temperatures were used, equally spaced between 0 and 1. The integral in Eq. (11) was estimated using the midpoint rule. Parameters of the method, such as δ in the proposal distribution or the length of the chain were the same for all dimensions and not optimized. The Markov chains were restarted repeatedly, and the length of the chains just made large enough to ensure convergence for $\mathcal{B}_{10}$. 95% confidence intervals for the volumes were computed from the values between restarts. In all cases, the exact value was located inside the confidence interval, and a relative error of better than 5% was achieved, if we define this as ratio of the half length of the 95% confidence interval and the exact value. Fig. 1 visualizes the good agreement of exact and Monte Carlo values with results varying over many orders of magnitude.

The estimation for $\mathcal{B}_{10}$ with an error of 5% in this implementation (not tuned for efficiency) took about 12 hours. An exact result is of much higher value of course, but for comparison, it is interesting to note that [8] reports that the total computation time needed to find the *exact* value of vol$\mathcal{B}_{10}$ on a workstation farm – scaled down to a single 1GHz processor – was almost 17 years.

Fig. 2 shows how the computation time required to get a relative error of 5% depends on the number of dimensions. A double logarithmic fit indicates a that the computation time increases with a power law d^ν with an exponent $\nu \approx 2.92$. This dependence has to be investigated for a larger class of polytopes and higher dimensions, but it provides some evidence that the algorithm proposed can compete with the best available algorithm that requires $O^*(d^4)$ oracle calls.

Table 1: MCMC volume estimation of the Birkhoff polytopes.

Polytope	Dimensions	Exact value	MCMC value	95% confidence interval	Computation time [sec]
$\mathcal{B}_5$	16	$1.40937 \cdot 10^{-4}$	$1.404 \cdot 10^{-4}$	$[1.398, 1.410] \cdot 10^{-4}$	31251
$\mathcal{B}_6$	25	$7.35257 \cdot 10^{-9}$	$7.357 \cdot 10^{-9}$	$[7.304, 7.411] \cdot 10^{-9}$	33679
$\mathcal{B}_7$	36	$5.63984 \cdot 10^{-15}$	$5.669 \cdot 10^{-15}$	$[5.604, 5.735] \cdot 10^{-15}$	35640
$\mathcal{B}_8$	49	$4.41855 \cdot 10^{-23}$	$4.463 \cdot 10^{-23}$	$[4.392, 4.535] \cdot 10^{-23}$	36822
$\mathcal{B}_9$	64	$2.59833 \cdot 10^{-33}$	$2.611 \cdot 10^{-33}$	$[2.549, 2.673] \cdot 10^{-33}$	42107
$\mathcal{B}_{10}$	81	$8.78201 \cdot 10^{-46}$	$8.860 \cdot 10^{-46}$	$[8.490, 9.247] \cdot 10^{-46}$	44649

4. Summary and conclusions

In this work, a direct Monte Carlo method for high-dimensional volume estimation based on the representation of the volume as a surface integral over the unit sphere was developed as well as an MCMC variant. The direct method is related to a method developed by Fok and Crevier, and already can give good results for interesting cases with a moderate number dimensions. For high dimensions, $d \gg 10$, the representation usually leads to a sharply peaked and multimodal integrand that requires the usage of importance sampling techniques. Markov Chain Monte Carlo methods for the computation of normalizing factors were used to derive an algorithm that also works in much higher dimensions. The methods were tested successfully for up to 81 dimensions, for an estimation of the volume of the Birkhoff polytope $\mathcal{B}_{10}$.

For the Birkhoff polytopes investigated, the computation time shows a $O(d^\nu)$-dependence on the number d of dimensions, with a comparatively small value $\nu \approx 3$. Since the best available algorithm for volume computation [2] has a running time of $O^*(d^4)$, it is interesting to study the behaviour of the method for a larger class of polytopes and higher dimensions. The method proposed can be further improved using more sophisticated path sampling techniques ([11], [12]) and adaptive MCMC methods ([14], [15]).

Acknowledgments

I thank Maik Kschischo, Claus Neidhardt and Matthias Beck for helpful discussions.

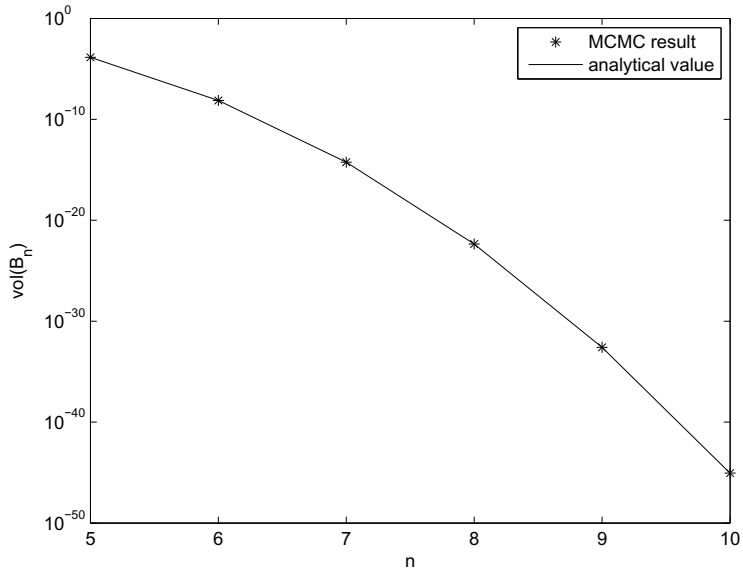

Figure 1: Volumes of the Birkhoff polytopes $\mathcal{B}_5 \ldots \mathcal{B}_{10}$.

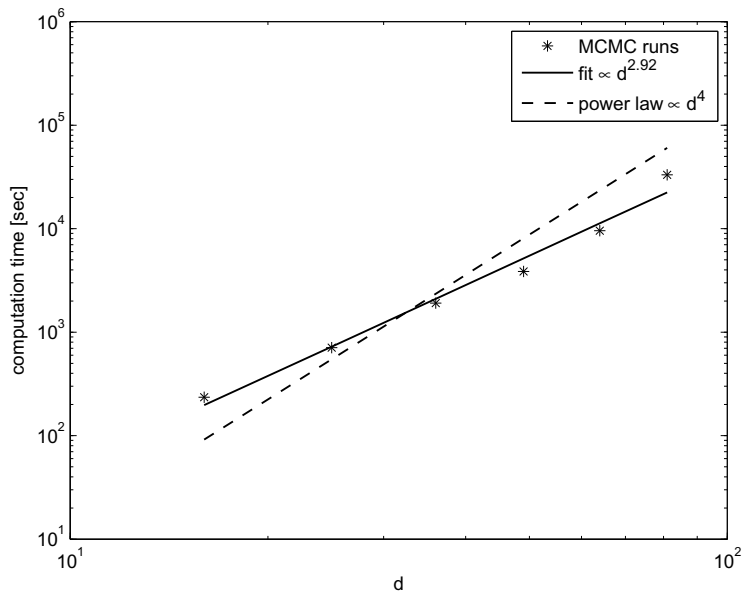

Figure 2: Dependence of the computation times required to achieve 5% accuracy using the MCMC method.

References

[1] M. Dyer *et al.*, 1991. JACM (38)1.
[2] L. Lovász and S. Vempala, 2006, J. Comput. Syst. Sci. (72) 392.
[3] D.S.K. Fok and D. Crevier, 1989, J. Statist. Comput. Simul.(31) 223.
[4] E. Almaas *et al.* , 2004, Global organization of metabolic fluxes in the bacterium Escher- ichia coli. Nature. 427 (6977), 839?84.
[5] S.J. Wiback *et al*, 2004, J. Theor. Biol. (228) 437.
[6] M. Kschischo, 2010, Eur. Phys. J. Special Topics (187) 255.
[7] M. Beck and D. Pixton, 2003, Discrete and Computational Geometry (30), 623.
[8] M. Beck and D. Pixton, `http://www.math.binghamton.edu/dennis/Papers/birkhoff.10.pdf`.
[9] M. Beck and D. Pixton, Volume of the Birkhoff polytope for n =1, ..., 10, `http://www.math.binghamton.edu/dennis/Birkhoff/volumes.html`.
[10] R. E. Kass and A. E. Raftery, Journal of the American Statistical Association, Vol. 90, No. 430. (Jun., 1995), pp. 773-795.
[11] A. Gelman and X.-L. Meng, 1998, Stat. Sci. (13), 163.
[12] I. Murray, Advances in Markov chain Monte Carlo methods, PhD thesis, 2007, Gatsby computational neuroscience unit, University College London, `http://homepages.inf.ed.ac.uk/imurray2/pub/07thesis/murray_thesis_2007.pdf`, and references therein.
[13] D. Gamerman and H.F. Lopes, 2006, Markov chain Monte Carlo: stochastic simulation for Bayesian Inference, 2nd edn. Chapman & Hall, London
[14] C. Andrieu and J. Thoms, 2008, Stat. Comput. (18), 343.
[15] H. Haario *et al.*, 2006, Stat. Comput. (16), 339.