Simulated Annealing in Convex Bodies and an $O^*(n^4)$ Volume Algorithm

László Lovász [1] Santosh Vempala [2]

Technical Report
MSR-TR-2003-31

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
http://www.research.microsoft.com

[1]Microsoft Research, One Microsoft Way, Redmond, WA 98052. lovasz@microsoft.com
[2]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139.

1 Introduction

Efficient volume computation in high dimension is an important question both theoretically and practically. The first polynomial time randomized algorithm to compute the volume of a convex body in $\mathbb{R}^n$ was given by Dyer, Frieze and Kannan in their pathbreaking paper [3]. A very high power of the dimension n (about 26) occurred in the running time bound of this algorithm, but subsequent improvements brought the exponent down to 5 [7]. In this paper, we further improve the running time to $O^*(n^4)$ (where the asterisk means that we suppress factors that are logarithmic in n). The algorithm uses $O^*(n)$ points for its computations, and in this sense it is nearly optimal, since any algorithm must use $\Omega(n)$ points.

The main ingredient of our algorithm is a method that can be viewed as a variation of *simulated annealing*. This method was briefly described in [6], but it was considered as a generalization of the volume computation algorithm rather than a tool for improvement.

Simulated annealing, introduced by Kirkpatrick et al. [5], is a general-purpose randomized search method for optimization. It does a random walk in the space of possible solutions, gradually adjusting a parameter called "temperature". At high temperature, the random walk converges fast to the uniform distribution over the whole space; as the temperature drops, the stationary distribution becomes more and more biased towards the optimal solutions. Simulated annealing often works well in practice, but it is notoriously difficult to obtain any theoretical guarantees for its performance.

To explain the connection between volume computation and simulated annealing, let us review the common structure of previous volume algorithms. All these algorithm reduce volume computation to sampling from a convex body, using the "Multi-Phase Monte-Carlo" technique. One constructs a sequence of convex bodies $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = k$, where K_0 is a body whose volume is easily computed, and one estimates the ratios $\mathrm{vol}(K_{i-1})/\mathrm{vol}(K_i)$ $(i = 1, \ldots, m)$ by generating sufficiently many independent uniformly distributed random points in K_i and counts what fraction of them falls in K_{i-1}. The generation of random points in K_i is done by some version of the Markov chain method (lattice walk, ball walk, hit-and-run), whose details can be ignored for the moment.

Of course, one would like to choose the number of phases, m, to be small. Any saving in the number of phases enters as its square in the running time: not only through the reduced number of iterations but also through the fact that we can allow larger errors in each phase, which means a smaller number of sample points are needed.

However, reducing the number of phases is constrained by the fact that in order to get a sufficiently good estimate for the ratio $\mathrm{vol}(K_{i-1})/\mathrm{vol}(K_i)$, one needs about $m\mathrm{vol}(K_i)/\mathrm{vol}(K_{i-1})$ random points. It follows that the ratios $\mathrm{vol}(K_i)/\mathrm{vol}(K_{i-1})$ must not be too large; since the volume ratio between $\mathrm{vol}(K)$ and $\mathrm{vol}(K_0)$ is $n^{\Omega(n)}$ in the worst case for any conceivable choice of K_0, it follows that m has to be $\Omega(n)$ just to keep the ratios $\mathrm{vol}(K_i)/\mathrm{vol}(K_{i-1})$ polynomial size. It turns out that the best choice is to keep these ratios bounded; this can be achieved e.g. if $K_0 = B$ is the unit ball and $K_i = K \cap (2^{i/n}B)$ for $i = 1, 2, \ldots, m = \Theta(n \log n)$. (After appropriate preprocessing, one can assume that $B \subseteq K \subseteq O(\sqrt{n})B$.) Reducing m any further (i.e., $o(n)$) appeared to be a fundamental hurdle.

On the other hand, volume computation is a special case of integration. Since the paper of Applegate and Kannan [1], the flexibility obtained by extending the problem to the integration of special kinds of functions (mostly logconcave) has been exploited in several papers. Mostly integration was used to dampen the boundary effects; we use it in a different way. Instead of a sequence of bodies, we construct a sequence of functions $f_0 \leq f_1 \leq \cdots \leq f_m$ that "connect" a function f_0 whose integral is easy to find to the characteristic function f_m of K. The ratios

$(\int f_{i-1})/(\int f_i)$ can be estimated by sampling from the distribution whose density function is proportional to f_i, and averaging the function f_{i-1}/f_i over the sample points.

If the f_i are characteristic functions of the K_i, then this is just the standard algorithm. The crucial gain comes from the fact that the number of sample points needed in each phase is smaller if the f_i are smooth. We use functions of the form $f(x) = e^{-x_0/T}$, where x_0 is the first coordinate of x (we'll come back to the preprocessing of K that is needed). For this choice, we'll only need $O^*(\sqrt{n})$ phases, and $O^*(\sqrt{n})$ sample points in each phase.

On two points this new approach brings in new difficulties. First, we have to sample from distributions that are not uniform over K. Various methods for sampling have been extended to logconcave distributions, and indeed our density functions are logconcave; but they do not satisfy any smoothness conditions, and so we have to use recent results [8, 9, 10] that give sampling algorithms with $O^*(n^3)$ steps (oracle calls) per sample point, without any smoothness assumption.

The other difficulty is that these sampling algorithms need a "warm start", i.e., they cannot be started from a fixed point but from a random point that is already almost uniformly distributed, in the sense that their density function (relative to the stationary distribution) is bounded. In the standard versions of the volume algorithm, this could be guaranteed by using the sample points generated in the preceding phase as starting points for the new phase. In our case this cannot be done, since these densities are not bounded. Instead, we use recent results from [10], which enable us to do efficient sampling if we only know that the L_2 norm of the starting density is bounded.

2 Outline of volume algorithm

2.1 Sampling

In our algorithm, we use as a black box a sampling algorithm (or *sampler* for short), which samples from a distribution supported on a convex body K, whose density is proportional to a given exponential function $e^{-a^T x}$; the algorithm needs a starting point $X \in K$.

Convex Body Sampler:

- *Input:* a convex body $K \in \mathbb{R}^n$, a vector $a \in \mathbb{R}^n$, a starting point $X \in K$, a warm start measure M, and an accuracy parameter $\varepsilon > 0$;
- *Output:* a point $Y \in K$.

A sampler we can use was given in [10], using an implementation of the hit-and-run algorithm. We consider the density function

$$f(x) = \frac{e^{-a^T x}}{\int_K e^{-a^T y}\, dy}$$

and the corresponding probability measure μ_f. We make the following assumptions about the data:

(A1) Every level set L of f contains a ball with radius $\Omega(\mu_f(L))$.

(A2) $\int_K f(x)|x|^2\, dx = O(n)$.

(A3) The starting point X is a random point from a distribution σ whose L_2-norm with respect to μ_f is at most M. This norm is defined as

$$\|\sigma\| = \int_K \frac{d\sigma}{d\mu_f} d\sigma = \int_K \left(\frac{d\sigma}{d\mu_f}\right)^2 d\mu_f.$$

Then the total variation distance of the output distribution from μ_f is less than ε. Furthermore, the number of calls on the membership oracle is

$$O(n^3 \ln^5 \frac{Mn}{\varepsilon^2}).$$

If M is polynomially bounded in n (as it will be in our case), this bound is $O^*(n^3)$. The number of other arithmetic operations is $O^*(n^5)$, on numbers with a polylogarithmic number of digits. As in all previous algorithms, it is a factor of $O^*(n^2)$ more than the oracle complexity.

We remark that (A1) and (A2) hold if f is in isotropic position. But it will be important for us that these weaker conditions are sufficient.

2.2 Rounding the body

We assume that K contains the unit ball B and is contained in the ball DB of radius D about the origin, where $D = O(\sqrt{n})$. This can be achieved by applying an appropriate affine transformation: we generate $O(n \log^2 n)$ approximately uniformly distributed random points in K, and bring these points (more exactly, the uniform measure on this finite set) into isotropic position. By results of Bourgain [2] and Rudelson [11], this brings the body into isotropic position. If we replace the body by its intersection with the ball DB, then we only loose a $o(1)$ fraction of the volume, and so we have achieved the well-rounded position.

The outline of this is as follows. Assume that we begin with a point that is the center of a unit ball B contained in K. We consider a sequence of convex bodies $K_i = K \cap 2^i B$. By the results of [10], we can get $O^*(n)$ samples from K_i in $O^*(n^4)$ steps (the first point in $O^*(n^4)$ steps and the remaining in $O^*(n^3)$ each). We use these to put K_i in near-isotropic position. The number of iterations is $O(\log D)$ and so the overall complexity is $O^*(n^4)$. In [7], the rounding was interlaced with volume computation. This can be done with the current algorithm also, but it will be conceptually easier to think of it as a preprocessing step. We don't go into more details in this abstract.

2.3 The pencil construction

Let K be the given body in $\mathbb{R}^n$ and $\varepsilon > 0$. Let C denote the cone in $\mathbb{R}^{n+1}$ defined by

$$C = \{x \in R^{n+1} \ : \ 2|x| \le x_0\}$$

(where $x = (x_0, x_1 \ldots, x_n)^\mathsf{T}$). We define a new convex body $K' \in \mathbb{R}^{n+1}$ as follows:

$$K' = \Big([0, 2D] \times K\Big) \cap C.$$

In other words, K' is an $(n+1)$-dimensional "pencil" whose cross-section is K, which is sharpened and its point is at the origin. Note that by the definition of D, the part of K' in the halfspace $x_0 \ge D$ is inside C and so it is a cylinder over K, while the part of K' in the halfspace $x_0 \le 1$

is a cone C_B over the unit ball. See Fig. 1 in the appendix for an illustration. It is trivial to implement a membership oracle for K'.

The sharpening took less than half of the volume of the pencil away. Hence if know the volume of K', it is easy to estimate the volume of K by generating $1/\varepsilon^2$ sample points from the uniform distribution on $[0, 2D] \times K$ and then counting how many of them fall into K'. Note that K' is also well-rounded (if we shift the origin to the point $(D, 0)$).

2.4 The multi-phase Monte-Carlo

For each real number $a > 0$, let

$$Z(a) = \int_{K'} e^{-ax_0}\, dx$$

where x_0 is the first coordinate of x. For $a \leq \varepsilon^2/D$, an easy computation shows that have $(1-\varepsilon)\mathrm{vol}(K') \leq Z(a) \leq \mathrm{vol}(K)$, so it suffices to compute $Z(a)$ for such an a. On the other hand, for $a \geq 2n$ the value of $Z(a)$ is essentially the same as the integral over the whole cone, which is easy to compute:

$$Z(a) \leq \int_C e^{-ax_0}\, dx = \int_0^\infty e^{-at} t^n \pi_n\, dt = n! \pi_n a^{-(n+1)}.$$

and

$$Z(a) \geq \int_{C_B} e^{-ax_0}\, dx = \int_0^1 e^{-at} t^n \pi_n\, dt > (1-\varepsilon) \int_0^\infty e^{-at} t^n \pi_n\, dt$$

by standard computation.

So if we select a sequence $a_0 > a_1 > \cdots > a_m$ for which $a_0 \geq 2n$ and $a_m \leq \varepsilon^2/D$, then we can estimate $\mathrm{vol}(K')$ by

$$Z(a_m) = Z(a_0) \prod_{i=0}^{T-1} \frac{Z(a_{i+1})}{Z(a_i)}.$$

Next we have to estimate the ratios

$$R_i = \frac{Z(a_{i+1})}{Z(a_i)}. \tag{1}$$

Let μ_i be the probability distribution over K' with density proportional to $e^{-a_i x_1}$, i.e.

$$\frac{d\mu_i(x)}{dx} = \frac{e^{-a_i x_1}}{Z(a_i)}.$$

Let X be a random sample point from μ_i, and let $Y = e^{(a_i - a_{i+1})X_0}$. It is easy to verify that Y has expectation R_i:

$$\begin{aligned} \mathsf{E}(Y) &= \int_{K'} e^{(a_i-a_{i+1})x_0}\, d\mu(x) = \int_{K'} e^{(a_i-a_{i+1})x_0} \frac{d\mu_i(x)}{dx}\, dx \\ &= \int_{K'} e^{(a_i-a_{i+1})x_0} \frac{e^{-a_i x_1}}{Z(a_i)}\, dx = \frac{1}{Z(a_i)} \int_{K'} e^{-a_{i+1}x_0}\, dx = \frac{Z(a_{i+1})}{Z(a_i)}. \end{aligned}$$

So to estimate the ratio R_i, we draw random samples $X^1, \ldots, X^k$ from μ_i, and compute the average

$$W_i = \frac{1}{m} \sum_{j=1}^{m} e^{(a_i - a_{i+1})(X^j)_0}.$$

Sample points from μ_0 are easy to get: select a random positive real number X_0 from the exponential distribution with density proportional to e^{-2nx}, and a uniform random point $(Y_1, \ldots, Y_n)$ from the unit ball B. If $X = (X_0, X_0Y_1, \ldots, X_0Y_n) \notin K'$, try again; else, return X.

In order to get appropriate sample points from μ_i $(i > 0)$, we have to make a simple affine transformation. Let $\gamma_i = \max(1, a_i/\sqrt{n})$, and

$$(T_i x)_j = \begin{cases} \gamma_i x_0 & \text{if } j = 0, \\ x_j & \text{otherwise.} \end{cases}$$

The algorithm shown in the box takes as input the dimension n of K, a sampling oracle for μ_i $(i = 1, \ldots, m)$, and an accuracy parameter ε. Its output Z is an estimate of the volume of K', correct to within a $1 \pm \frac{\varepsilon}{2}$ factor, with high probability.

Volume algorithm:

(V1) Set $m = 2\lceil \sqrt{n} \ln \frac{n}{\varepsilon} \rceil$, $\quad k = \frac{8}{\varepsilon^2}\sqrt{n} \ln \frac{n}{\varepsilon}$, $\quad \delta = n^{-10}$ $\quad$ and

$$a_i = 2n(1 - \frac{1}{\sqrt{n}})^i \quad \text{for } i = 1, \ldots, m.$$

(V2) For $i = 0, 1, \ldots, m$, do the following.

— Run the sampler k times for convex body T_iK, with exponential function $e^{-a_i x_0/\gamma_i}$, error parameter δ, and (for $i > 0$) starting points $T_iX_{i-1}^1, \ldots T_iX_{i-1}^k$. Apply T_i^{-1} to the resulting points to get points $X_i^1, \ldots, X_i^k$.

— Using these points, compute

$$W_i = \frac{1}{k} \sum_{j=1}^{k} e^{(a_i - a_{i+1})(X_i^j)_0}. \tag{2}$$

(V3) Return

$$Z = n! \pi_n a^{-(n+1)} W_1 \ldots W_m$$

as the estimate of the volume of K'.

The sequence of sample points $(X_0^j, X_2^j, \ldots, X_m^j)$ for a fixed j is called a *thread*. Note that the threads are independent.

The analysis of the algorithm will need the verification of a number facts. Specifically, we are going to show:

1. The variance of the function $e^{(a_i - a_{i+1})x_0}$ relative to the distribution μ_i is small enough so that k sample points suffice to estimate its mean (Lemma 4.1).

2. Random samples from one phase provide a warm start for the next phase (Lemma 4.4).

3. The convex body T_iK and exponential function $e^{-a_i x_0/\gamma_i}$ satisfy (A1) and (A2) (Lemmas 4.5, 4.6).

4. The overall complexity is

$$O^*(\sqrt{n}) \text{ phases } \times O^*(\sqrt{n}) \text{ samples per phase } \times O^*(n^3) \text{ steps per sample } = O^*(n^4).$$

To be more precise,

Theorem 2.1 *The volume of a convex body K, given by a membership oracle, can be approximated to within a relative error of ε with probability $1-\delta$ in*

$$O(\frac{n^4}{\varepsilon^2}\log^7\frac{n}{\varepsilon\delta}) = O^*(n^4)$$

oracle calls.

This will need some technical preliminaries, collected in the next section.

3 Preliminaries

3.1 Logconcavity

The following lemma about logconcave functions will play a key role.

Lemma 3.1 *Let $K \subseteq \mathbb{R}^n$ be a convex body and $f : K \to \mathbb{R}$, a logconcave function. For $a > 0$, define*

$$Z(a) = \int_K f(ax)\,dx.$$

Then $a^n Z(a)$ is a logconcave function of a.

Proof. Let

$$G(x,t) = \begin{cases} 1 & \text{if } t > 0 \text{ and } (1/t)x \in K, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that $G(x,t)$ is logconcave, and so the function

$$F(x,t) = f(x)G(x,t)$$

is also logconcave. It follows that its marginal in t is a logconcave function of t. But this marginal is just

$$\int_{\mathbb{R}^n} f(x)G(x,t)\,dx = t^n \int_K f(tx)\,dx.$$

□

3.2 Probability

Two random variables X, Y will be called *μ-independent* ($0 < \mu < 1$) if for any two sets A, B in their ranges,

$$\big|\mathsf{P}(X \in A,\ Y \in B) - \mathsf{P}(X \in A)\mathsf{P}(Y \in B)\big| \le \mu.$$

Several basic properties of this notion are collected in the Appendix.

4 Analysis of volume algorithm

4.1 Variance

We begin by bounding the variance of our sampling estimate for R_i.

Lemma 4.1 *Let X be a random sample from $d\mu_i$, $h \in \mathbb{Z}_+$ and let*

$$Y = e^{(a_i - a_{i+1})X_0}.$$

Then

$$\mathsf{E}(Y^2) \le eE(Y)^2.$$

Proof. For notational convenience, let $a = a_i$ and $b = a_{i+1} - a_i$. We have

$$\mathsf{E}(Y) = \frac{\int_{K'} e^{-(a+b)x_0}\,dx}{\int_{K'} e^{-ax_0}\,dx}$$

and

$$\mathsf{E}(Y^2) = \frac{\int_{K'} e^{-(a+2b)x_0}\,dx}{\int_{K'} e^{-ax_0}\,dx}.$$

By Lemma 3.1 the value $a^{n+1}\int_{K'} e^{-ax_0}\,dx$ is a logconcave function of a, and so

$$\int_K e^{-ax_0}\,dx \int_K e^{-(a+2b)x_0}\,dx \le \left(\frac{(a+b)^2}{a(a+2b)}\right)^{n+1} \left(\int_K e^{-(a+b)x_0}\,dx\right)^2.$$

Since we have $b = a/\sqrt{n}$, the coefficient on the right hand side is

$$\left(\frac{(a+b)^2}{a(a+2b)}\right)^{n+1} = \left(1 + \frac{1}{n+2\sqrt{n}}\right)^{n+1} < e.$$

□

This lemma proves that k sample points are enough to estimate R_i by (2) with relative error ε/m, with probability $1 - 1/n^{10}$.

4.2 Divine intervention

Our goal is to prove

Lemma 4.2 *With probability $1 - o(1)$,*

$$(1-\varepsilon)R_1 \dots R_m \le W_1 \dots W_m \le (1+\varepsilon)R_1 \dots R_m.$$

Proof. To analyze the algorithm we use the "divine intervention" method. The distribution of the random point X_i^j is *approximately* μ_i. We construct modified random variables $\overline{X}_i^j$ ($i = 0, \dots, m$, $j = 1, \dots, k$) whose distribution is *exactly* μ_i as follows. Fix j. We define $\overline{X}_0^j = X_0^j$. Assuming that $\overline{X}_i^j$ is defined, let Z be the random point returned by the sampler S_i when it is started from $\overline{X}_i^j$. Let ν denote the total variation distance of the distribution of Z from the distribution μ_{i+1}. By the specification of the sampler, $nu \le \delta$. Then we define $\overline{X}_{i+1}^j$ as a random

variable with distribution μ_{i+1} such that $\mathsf{P}(Z = X_{i+1}^j) = 1 - \nu$. The construction is carried out independently for each thread, so that the modified threads $(\overline{X}_0^j, \overline{X}_1^j, \ldots, \overline{X}_m^j)$ are independent.

Assume that $X_i^j = \overline{X}_i^j$. Then $X_{i+1}^j = Z$, and so $\overline{X}_{i+1}^j = X_{i+1}^j$ with probability at least $1 - \delta$. It follows by induction that $X_i^j = \overline{X}_i^j$ is at least $1 - j\delta$, and hence

$$\mathsf{P}(X_i^j = \overline{X}_i^j \text{ for all } j) \geq 1 - km\delta. \tag{3}$$

[It would be nice to use "divine intervention" to achieve that the X_i^j in one of the threads be independent, but this does not work (for this, the sampler would have to work with a cold start, which would take too long). We'll have to estimate the dependence between consecutive phases carefully.]

Since the random variables $\overline{X}_i^j$ have the "right" distribution μ_i, we have

$$\mathsf{E}(e^{a_i - a_{i+1}}(\overline{X}_i^j)_0) = R_i.$$

Let

$$\overline{W}_i = \frac{1}{k} \sum_{j=1}^{k} e^{a_i - a_{i+1}(\overline{X}_i^j)_0}.$$

Then $\mathsf{E}(\overline{W}_i) = R_i$, and by Lemma 4.1, $\mathsf{E}(\overline{W}_i^2) \leq (1 + \frac{e-1}{k})R_i^2$.

Now consider the product $\overline{Z} = \overline{W}_1 \overline{W}_2 \ldots \overline{W}_m$. If we had independence between successive phases, then we would have

$$\mathsf{E}(\prod_{i=1}^{m} \overline{W}_i) = \prod_{i=1}^{m} R_i$$

and

$$\mathsf{E}(\prod_{i=1}^{T} \overline{W}_i^2) \leq (1 + \frac{e-1}{k})^m \prod_{i=1}^{m} R_i^2.$$

This would imply $Var(\overline{Z}) \leq (1 + \frac{e-1}{k})^m - 1$, and we could easily show that with probability at least $3/4$, $\overline{Z}$ is within a factor of $(1 \pm \varepsilon)$ to the volume of K'. Since $Z = \overline{Z}$ with probability $1 - o(1)$, it follows that with probability at least $3/4 - o(1)$, $\overline{Z}$ is within a factor of $(1 \pm \varepsilon)$ to the volume of K'.

Unfortunately, since we are using the sample points from each phase as the starting points for the next, the random variables W_i are only approximately independent as shown in the next Lemma.

Lemma 4.3 (a) *For every phase $0 \leq i < m$ and every thread j, the random variables X_i^j and X_{i+1}^j are δ-independent, and the random variables $\overline{X}_i^j$ and $\overline{X}_{i+1}^j$ are (3δ)-independent.*

(b) *For every phase $0 \leq i < m$ and every thread $1 \leq j \leq k$, the random variables $(X_0^j, \ldots, X_i^j)$ and X_{i+1}^j are (3δ)-independent.*

(c) *For every phase $0 \leq i < m$, the random variables $\overline{W}_1 \ldots \overline{W}_i$ and $\overline{W}_{i+1}$ are $(3k\delta)$-independent.*

The variables $\overline{W}_1 \ldots \overline{W}_i$ and $\overline{W}_{i+1}$ are not bounded, so we cannot apply Lemma 6.2 directly. So we define another set of random variables

$$V_i = \min\{\overline{W}_i, \frac{1}{\sqrt{n\mu}} \mathsf{E}(\overline{W}_i)\}$$

where $\mu = 3k\delta$. The rest of the proof is given in the appendix. □

4.3 Warm start

The next lemma shows that samples from one phase provide a warm start for the next phase.

Lemma 4.4 *The L_2-norm of μ_i with respect to μ_{i+1} is at most 4.*

Proof. Let X be a random sample from μ_i. Then we have to prove that

$$\mathsf{E}\left[\frac{d\mu_i(X)}{d\mu_{i+1}(X)}\right] \le 4.$$

Indeed, using Lemma 3.1

$$\begin{aligned}
\mathsf{E}\left[\frac{d\mu_i(X)}{d\mu_{i+1}(X)}\right] &= \frac{\int_K e^{(a_{i+1}-a_i)X_1}e^{-a_iX_1}dx}{\int_K e^{-a_ix_1}dx}\frac{\int_K e^{-a_{i+1}x_1}dx}{\int_K e^{-a_ix_1}dx} \\
&= \frac{Z(2a_i-a_{i+1})Z(a_{i+1})}{Z(a_i)Z(a_i)} \\
&\le \left(\frac{(2a_i)^2}{4a_{i+1}(2a_i-a_{i+1})}\right)^n \\
&= \left(\frac{1}{(1-\frac{1}{\sqrt{n}})(1+\frac{1}{\sqrt{n}})}\right)^n < 4.
\end{aligned}$$

□

4.4 Rounding

For $s \ge 0$, let $K_s = \{x \in K' : \ x_0 \le s\}$. Note that these sets are exactly the level sets of the functions $f_i = e^{-a_ix_0}$.

Lemma 4.5 *Let $c = \mu_i(K_s)$. Then T_iK_s contains a ball with radius $c/10$.*

Proof. Let X be a random point from μ_i, and let X_0 be its first coordinate. We denote by F the density function of X_0.

The intersection of the hyperplane $x_0 = s$ with K' contains a ball with radius $\min(1, s)$. Hence the body T_iK' contains a cone with height $\gamma_i s \ge s$ over this ball. If we show that $s > c/4$, then it follows by simple geometry that K_s contains a ball with radius $c/12$.

We may assume that $s < 1/4$. Let $F(t)$ denote the density function of X_0. This function is proportional to $t^n e^{-a_it}$ for $t < 1$. Using that $a_i \le 2n$, it follows that $F(t)$ is monotone increasing for $t \le 1/2$, and so its value is at least $F(s)$ between $1/4$ and $1/2$. Thus we have

$$c = \frac{\int_0^s F(t)\,dt}{\int_0^{2D} F(t)\,dt} \le \frac{sF(s)}{(1/4)F(s)} = 4s.$$

□

Lemma 4.6 *If X is a random point from the distribution μ_i, then $\mathsf{E}(|T_iX|^2) \le 5D^2$.*

Proof. Let X be a random point from μ_i, and let $Y = T_i X$. First we estimate the expectation of Y_0^2. If $a_i \le \sqrt{n}$, then $\gamma_i = 1$ and $Y = X$, so $|Y_0| \le 2D$, so $\mathsf{E}(Y_0^2) \le 4D^2$.

Let $a_i > \sqrt{n}$. Let Z be a random point from the distribution over the whole cone C with density function proportional to $e^{-a_i x_0}$. Then

$$\mathsf{E}(X_0^2) \le \mathsf{E}(Z_0^2) = \frac{\int t^{n+2} e^{-at}\, dt}{\int t^n e^{-at}\, dt} = \frac{(n+1)(n+2)}{a^2},$$

and hence

$$\mathsf{E}(Y_0^2) = \gamma_i^2 \mathsf{E}(X_0^2) \le \frac{a^2}{n} \frac{(n+1)(n+2)}{a^2} = \frac{(n+1)(n+2)}{n} < D^2.$$

The expectation of $Y_1^2 + \cdots + Y_n^2$, conditional on any $X_0 = t$, is at most D^2, since K is well-rounded. This proves the lemma. □

5 Concluding remarks

1. If we view the sampler as a blackbox, then the number of calls to the sampler is $O^*(n)$, and this is the total number of points used to estimate the volume. In this sense, the algorithm is nearly optimal, since any algorithm has to examine $\Omega(n)$ points.

2. The same "simulated annealing" technique can be used to maximize a linear objective function over a convex body. Indeed, a randomized algorithm for this maximization problem can be thought of as a way of generating a sample from a distribution that is heavily biased in the direction of the objective function. Starting from the uniform distribution, and gradually increasing the bias, do we get an analyzable version of the simulated annealing for optimization?

3. It is a natural next step to extend this method to integration of logconcave functions. The fundamental lemma 3.1 can be extended to this case, but certain technical results we used from [10] are still not known for the general case. We believe that these difficulties can be overcome, and one can design an $O^*(n^4)$ integration algorithm for logconcave functions (after appropriate preprocessing).

4. How far can the exponent in the volume algorithm be reduced? There is one possible further improvement on the horizon. This depends on a difficult open problem in convex geometry, the "Slicing Conjecture". If this conjecture is true, then the mixing time of the hit-and-run walk in a convex body in isotropic position could be reduced to $O^*(n^2)$, which, when combined with ideas of this paper, could perhaps lead to an $O^*(n^3)$ volume algorithm. But besides the mixing time, a number of further problems concerning achieving isotropic position would have to be solved.

Acknowledgement. We are deeply grateful to Adam Kalai for many useful comments and ideas and for his help in simplifying the proof of Lemma 3.1.

References

[1] D. Applegate and R. Kannan: Sampling and integration of near log-concave functions, *Proc. 23th ACM STOC* (1990), 156–163.

[2] J. Bourgain: Random points in isotropic convex sets, in: *Convex geometric analysis* (Berkeley, CA, 1996), 53–58, Math. Sci. Res. Inst. Publ., 34, Cambridge Univ. Press, 1999.

[3] M. Dyer, A. Frieze and R. Kannan: A random polynomial-time algorithm for approximating the volume of convex bodies, *J. Assoc. Comput. Mach.* **38** (1991), 1–17.

[4] R. Kannan, L. Lovász and M. Simonovits: Isoperimetric problems for convex bodies and a localization lemma. *J. Discr. Comput. Geom.* **13** (1995) 541–559.

[5] S. Kirkpatrick, C.D. Gelatt Jr, and M.P. Vecchi: Optimization by Simulated Annealing, *Science* **220** (1983), 671—680.

[6] L. Lovász and M. Simonovits: On the randomized complexity of volume and diameter, *Proc. 33rd IEEE Annual Symp. on Found. of Comp. Sci.* (1992), 482–491.

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

[8] L. Lovász and S. Vempala: The geometry of logconcave functions and an $O^*(n^3)$ sampling algorithm, Microsoft Research Tech. Rep. MSR-TR-2003-04. Available at: http://www-math.mit.edu/~vempala/papers/logcon-ball.ps

[9] L. Lovász and S. Vempala: Hit-and-run is fast and fun, Microsoft Research Tech. Rep.MSR-TR-2003-05. Available at http://www-math.mit.edu/~vempala/papers/logcon-hitrun.ps

[10] L. Lovász and S. Vempala: Where to start a geometric random walk. Available at: http://www-math.mit.edu/~vempala/papers/start.ps

[11] M. Rudelson: Random vectors in the isotropic position, *J. Funct. Anal.* **164** (1999), 60–72.

Appendix

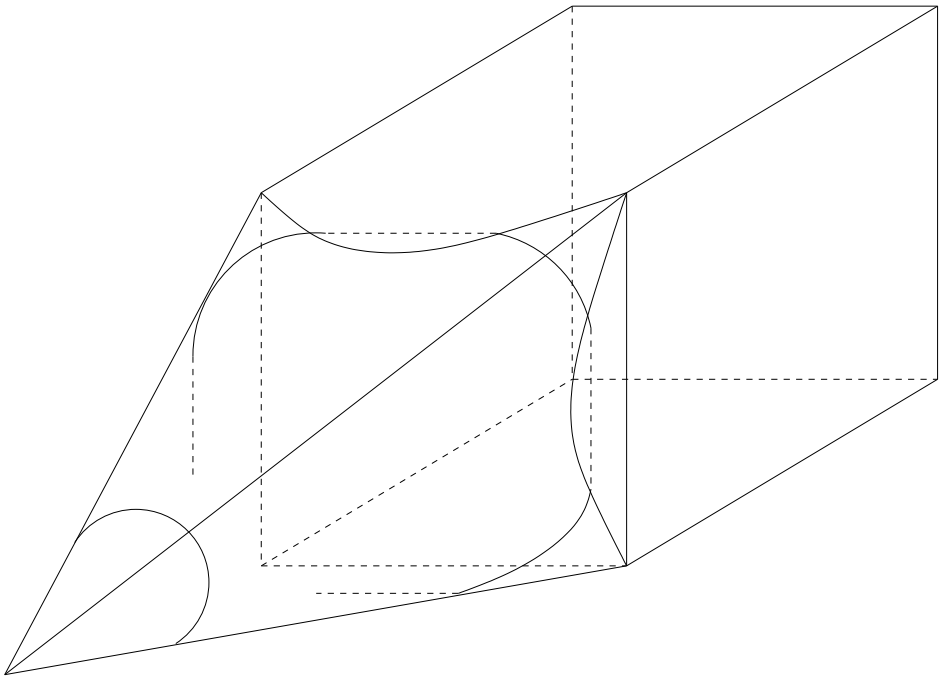

Figure 1: The pencil construction when K is a square.

The following are some basic properties of μ-independence.

Lemma 6.1 *If X and Y are μ-independent, and f, g are two measurable functions, then $f(X)$ and $g(Y)$ are also μ-independent.*

Proof. We have

$$\begin{aligned}&\big|\mathsf{P}(f(X) \in A,\ g(Y) \in B) - \mathsf{P}(f(X) \in A)\mathsf{P}(f(Y) \in B)\big| \\ &= \big|\mathsf{P}(X \in f^{-1}(A),\ Y \in g^{-1}(B)) - \mathsf{P}(X \in f^{-1}(A))P(Y \in g^{-1}(B))\big| \\ &\leq \mu.\end{aligned}$$

□

Another useful fact is the identity

$$\big|\mathsf{P}(X \in A,\ Y \in B) - \mathsf{P}(X \in A)\mathsf{P}(Y \in B)\big| = \big|\mathsf{P}(X \in \overline{A},\ Y \in B) - \mathsf{P}(X \in \overline{A})\mathsf{P}(Y \in B)\big|, \quad (4)$$

which implies that to check μ-independence, it suffices to consider sets A, B with $\mathsf{P}(X \in A) \geq 1/2$ and $\mathsf{P}(Y \in B) \geq 1/2$.

Lemma 6.2 *Let $X, Y \geq 0$ be μ-independent random variables such that $|X| \leq a$ and $|Y| \leq b$. Then*

$$\big|\mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)\big| \leq \mu ab.$$

This is a variation of a lemma in [7]), and the proof is the same.

Lemma 6.3 *Let X, Y, X', Y' be random variables, and assume that (a) the pair (X, Y) is independent from the pair (X', Y'), (b) X is μ-independent from Y and (c) X' is μ'-independent from Y'. Then the pair (X, X') is $(\mu + \mu')$-independent from the pair (Y, Y').*

Proof. Let R, R', S, S' be the range of $X, Y, X'Y'$, respectively, and let $A \subseteq R \times R'$, $B \subseteq S \times S'$ be measurable sets. We want to show that

$$\big|\mathsf{P}((X, X') \in A,\ (Y, Y') \in B) - \mathsf{P}((X, X') \in A)\mathsf{P}((Y, Y') \in B)\big| \leq \mu + \mu'. \tag{5}$$

For $r \in R$ and $s \in S$, let $A_r = \{r' \in \mathbb{R}' : (r, r') \in A\}$, $B_s = \{s' \in S' : (s, s') \in B\}$, $f(r) = \mathsf{P}(X' \in A_r)$, $g(s) = \mathsf{P}(Y' \in B_s)$ and $h(r, s) = \mathsf{P}(X' \in A_r, Y' \in B_s)$. Then

$$\mathsf{P}((X, X') \in A) = \mathsf{E}(f(X)), \quad \mathsf{P}((Y, Y') \in B) = \mathsf{E}(g(Y))$$

and

$$\mathsf{P}((X, X') \in A,\ (Y, Y') \in B) = \mathsf{E}(h(X, Y))$$

(here we use that (X, Y) is independent of (X', Y'). We can write the left hand side of (5) as

$$\begin{aligned} &\mathsf{E}(h(X, Y)) - \mathsf{E}(f(X))\mathsf{E}(g(Y)) \\ &= \Big[\mathsf{E}(h(X, Y) - f(X)g(Y))\Big] + \Big[\mathsf{E}(f(X)g(Y)) - \mathsf{E}(f(X))\mathsf{E}(g(Y))\Big] \end{aligned} \tag{6}$$

By assumption,

$$|h(r, s) - f(r)g(s)| = \big|\mathsf{P}(X' \in A_r, Y' \in B_s) - \mathsf{P}(X' \in A_r)\mathsf{P}(Y' \in B_s)\big| \leq \mu'$$

for every r and s, and hence the first term on the right hand side in (6) is at most μ' in absolute value. The second term is at most μ by Lemma 6.2. This proves (5). □

Lemma 6.4 *Let $X_0, X_1, \ldots,$ be a Markov chain, and assume that for some $i > 0$, X_{i+1} is μ-independent from X_i. Then X_{i+1} is μ-independent from $(X_0, \ldots, X_i)$.*

Proof. Let S_i be the range of X_i, and let $A \subseteq S_0 \times \cdots \times S_i$, $B \subseteq S_{i+1}$. We want to prove that

$$\big|\mathsf{P}((X_0, \ldots, X_i) \in A,\ X_{i+1} \in B) - \mathsf{P}((X_0, \ldots, X_i) \in A)\mathsf{P}(X_{i+1} \in B)\big| \leq \mu. \tag{7}$$

For $r \in S_i$, let $f(r) = \mathsf{P}((X_0, \ldots, X_{i-1}, r) \in A)$. Let g denote the characteristic function of B. Then

$$\mathsf{P}((X_0, \ldots, X_{i-1}, X_i) \in A) = \mathsf{E}(f(X_i)), \quad \text{and} \quad \mathsf{P}(X_{i+1} \in B) = \mathsf{E}(g(X_{i+1})).$$

For every $r \in S_i$,

$$\begin{aligned} &\mathsf{P}((X_0, \ldots, X_{i-1}, r) \in A, X_{i+1} \in B) \\ &= \mathsf{P}((X_0, \ldots, X_{i-1}, r) \in A)\mathsf{P}(X_{i+1} \in B \mid X_i = r) \\ &= f(r)\mathsf{E}(g(X_{i+1}) \mid X_i = r) = \mathsf{E}(f(r)g(X_{i+1}) \mid X_i = r). \end{aligned}$$

by the Markov property, and so

$$\mathsf{P}((X_0, \ldots, X_{i-1}, X_i) \in A, X_{i+1} \in B) = \mathsf{E}(f(X_i)g(X_{i+1})).$$

So (7) follows from Lemma 6.2 again. □

We need another simple fact from probability:

Lemma 6.5 *Let $X \geq 0$ be a random variable, $a > 0$, and $X' = \min(X, a)$. Then*

$$\mathsf{E}(X') \geq \mathsf{E}(X) - \frac{\mathsf{E}(X^2)}{4a}.$$

Proof. Let $X'' = X - X'$. Note that $X'X'' = aX''$ (if $X'' \neq 0$ then $X' = a$). Using this,

$$\mathsf{E}(X^2) = \mathsf{E}((X' + X'')^2) \leq 4\mathsf{E}(X'X'') = 4a\mathsf{E}(X''),$$

which implies the assertion. □

Proof. [of Lemma 4.3] (a) Let $A, B \subseteq K$; we claim that

$$\left|\mathsf{P}(X_i^j \in A, X_{i+1}^j \in B) - \mathsf{P}(X_i^j \in A)\mathsf{P}(X_{i+1}^j \in B)\right| \leq \delta. \tag{8}$$

By the remark after (4), we may assume that $\mu_i(A) \geq 1/2$. Let μ_i' be the restriction of μ_i to A, scaled to be a probability measure. Then $\mu_i' \leq 2\mu_i$, and it is easy to check that $\chi^2(\mu_i', \mu_{i+1}) \leq 4\chi^2(\mu_i, \mu_{i+1}) + 3$. Hence by the basic property of the sampler S_i,

$$\left|\mathsf{P}(X_{i+1}^j \in B \mid X_i^j \in A) - \mathsf{P}(X_{i+1}^j \in B)\right| \leq \delta,$$

and so (8) holds. The second assertion is immediate, since putting a bar above the variables changes the probabilities in the condition by at most δ.

(b) Follows from Lemma 6.4.

(c) Follows from Lemma 6.3. □

Proof. (of Lemma 4.2.) Let

$$V_i = \min\{\overline{W}_i, \frac{1}{\sqrt{n\mu}}\mathsf{E}(\overline{W}_i)\}$$

where $\mu = 3k\delta$. By Lemma 6.5, we have

$$\mathsf{E}(V_i) \geq \mathsf{E}(W_i) - \sqrt{n\mu}\frac{\mathsf{E}(W_i^2)}{\mathsf{E}(W_i)} \geq (1 - 2\sqrt{n\mu})\mathsf{E}(W_i). \tag{9}$$

Define recursively

$$U_{i+1} = \min\{U_iV_{i+1}, \frac{1}{\sqrt{n\mu}}\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})\}.$$

By Lemma 6.1, the random variables U_i and V_{i+1} are μ-independent, so it follows by Lemma 6.2 that

$$|\mathsf{E}(U_iV_{i+1}) - \mathsf{E}(U_i)\mathsf{E}(V_{i+1})| \leq \frac{1}{n}\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1}). \tag{10}$$

Next we show that

$$\mathsf{E}(U_i) \leq \left(1 + \frac{i}{n}\right)\mathsf{E}(V_1)\ldots\mathsf{E}(V_i). \tag{11}$$

This is easy by induction:

$$\mathsf{E}(U_{i+1}) \leq \mathsf{E}(U_iV_{i+1}) \leq \mathsf{E}(U_i)\mathsf{E}(V_{i+1}) + \frac{1}{n}\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})$$

(by (10))

$$\leq \left(1 + \frac{i+1}{n}\right)\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})$$

(using the induction hypothesis). A similar argument shows that

$$\mathsf{E}(U_i^2) \le \left(1+\frac{i}{n}\right)\mathsf{E}(V_1^2)\dots\mathsf{E}(V_i^2). \tag{12}$$

and

$$\mathsf{E}(U_i^2V_{i+1}^2) \le \left(1+\frac{i}{n}\right)\mathsf{E}(V_1^2)\dots\mathsf{E}(V_{i+1}^2). \tag{13}$$

Next we bound $\mathsf{E}(U_{i+1})$ from below. Using Lemma 6.5 and inequality (13), we get

$$\begin{aligned}\mathsf{E}(U_{i+1}) &\ge \mathsf{E}(U_iW_{i+1}) - \sqrt{n\mu}\frac{\mathsf{E}(U_i^2V_{i+1}^2)}{4\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1})}\\ &\ge \mathsf{E}(U_iW_{i+1}) - \sqrt{n\mu}\left(1+\frac{i}{n}\right)\frac{\mathsf{E}(V_1^2)\dots\mathsf{E}(V_{i+1}^2)}{4\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1})}.\end{aligned}$$

Here, using (9),

$$\mathsf{E}(V_i^2) \le \mathsf{E}(W_i^2) \le \left(1+\frac{e-1}{k}\right)\mathsf{E}(W_i)^2 \le \left(1+\frac{e-1}{k}\right)\frac{1}{1-2\sqrt{n\mu}}\mathsf{E}(V_i)^2 < \left(1+\frac{4}{k}\right)\mathsf{E}(V_i)^2),$$

so we get that

$$\mathsf{E}(U_{i+1}) \ge \mathsf{E}(U_iV_{i+1}) - \sqrt{n\mu}\left(1+\frac{i}{n}\right)\left(1+\frac{4}{k}\right)^i\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1}) \tag{14}$$

$$\ge \mathsf{E}(U_iV_{i+1}) - \frac{1}{n}\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1}). \tag{15}$$

We use (10) to estimate the first term:

$$\mathsf{E}(U_iV_{i+1}) \ge \mathsf{E}(U_i)\mathsf{E}(V_{i+1}) - \frac{1}{n}\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1}).$$

Thus

$$\mathsf{E}(U_{i+1}) \ge \mathsf{E}(U_i)\mathsf{E}(V_{i+1}) - \frac{2}{n}\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1}),$$

and hence by induction

$$\mathsf{E}(U_{i+1}) \ge \mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1}) - \frac{2i}{n}\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1}).$$

In particular, it follows that

$$\mathsf{E}(U_m) = (1+o(1))\mathsf{E}(V_1)\dots\mathsf{E}(V_m) = (1+o(1))\mathsf{E}(\overline{W}_1)\dots\mathsf{E}(\overline{W}_m).$$

By (12),

$$\mathsf{E}(U_m^2) = (1+o(1))\mathsf{E}(U_m)^2,$$

and hence U_m is close to its expectation with large probability.

Furthermore, using Markov's inequality,

$$\begin{aligned}\mathsf{P}(U_{i+1} \ne U_iV_{i+1}) &= \mathsf{P}\left(U_iV_{i+1} > \frac{1}{\sqrt{n\mu}}\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1})\right)\\ &\le \sqrt{n\mu}\frac{\mathsf{E}(U_iV_{i+1})}{\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1})} \le \sqrt{n\mu}\left(1+\frac{i+1}{n}\right) < 2\sqrt{n\mu},\end{aligned}$$

and similarly

$$\mathsf{P}(V_i \neq \overline{W}_i) \leq \sqrt{n\mu}.$$

So with probability at least $1 - 3m\sqrt{n\mu}$, we have $U_m = \overline{W}_1 \dots \overline{W}_m$. Furthermore, by (3), we have $\overline{W}_1 \dots \overline{W}_m = W_1 \dots W_m$ with probability at least $1 - km\delta$.

So $W_1 \dots W_m$ is close to $\mathsf{E}(W_1) \dots \mathsf{E}(W_m) = R_1 \dots R_m$ with large probability. This completes the proof of Lemma 4.2. □