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

László Lovász
Microsoft Research,
One Microsoft Way, Redmond, WA 98052.
lovasz@microsoft.com

Santosh Vempala*
Department of Mathematics,
MIT, Cambridge, MA 02139.
vempala@math.mit.edu

June 10, 2004

Abstract

We present a new algorithm for computing the volume of a convex body in $\mathbb{R}^n$. The main ingredient of the algorithm is a "morphing" technique that can be viewed as a variant of simulated annealing. Its complexity is $O^*(n^4)$, improving on the previous best algorithm by a factor of n.

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 [6]. This result is quite surprising, given that no deterministic polynomial-time algorithm can approximate the volume to within a factor that is exponential in n [7, 2]. A very high power of the dimension n (about 23) occurred in the complexity bound of the algorithm, but subsequent improvements [13, 1, 12, 5, 14, 15, 9] brought the exponent down to 5. In this paper, we further improve the running time to $O^*(n^4)$ (where the asterisk indicates that the dependence on the error parameter and on logarithmic factors in n is not shown).

The main ingredient of our algorithm is a method that can be viewed as a variation of *simulated annealing*. Introduced by Kirkpatrick et al. [10], simulated annealing 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 algorithms 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

*Supported by NSF Career award CCR-9875024 and a Sloan foundation fellowship.

volume is easily computed, and one estimates the ratios $\text{vol}(K_{i-1})/\text{vol}(K_i)$ $(i = 1, \dots, m)$ 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 $\text{vol}(K_{i-1})/\text{vol}(K_i)$, one needs about $m\text{vol}(K_i)/\text{vol}(K_{i-1})$ random points. It follows that the ratios $\text{vol}(K_i)/\text{vol}(K_{i-1})$ must not be too large; since the volume ratio between $\text{vol}(K)$ and $\text{vol}(K_0)$ 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 $\text{vol}(K_i)/\text{vol}(K_{i-1})$ 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 $i = 1, 2, \dots, m = \Theta(n \log n)$. (After appropriate preprocessing, one can assume that $B \subseteq K \subseteq O(\sqrt{n})B$; we will discuss such "roundings" in more detail.) Reducing m any further (i.e., $o(n)$) appears 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 boundary effects; we use it in a different way. Instead of a sequence of bodies, we construct a sequence of functions $f_0 \le f_1 \le \cdots \le f_m$ 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. (This method was briefly described in [14], but it was considered as a generalization of the volume computation algorithm rather than a tool for improvement.)

If the functions f_i are characteristic functions of the convex bodies 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. Thus, we get samples from a density proportional to $e^{-x_0/T}$ with monotonically changing values of T, similar to simulated annealing (where T is called the "temperature").

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 [16, 17, 18] 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 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 the hit-and-run random walk which has a much milder dependence on starting density. In [18], it was shown that efficient sampling by this walk requires only that the L_2 norm of the starting density is (polynomially) bounded.

The main result of the paper can be stated precisely as follows.

Theorem 1.1 *The volume of a convex body K, given by a membership oracle, and a parameter R such that $B \subseteq K \subseteq RB$, can be approximated to within a relative error of ε with probability $1 - \delta$ using*

$$O\left(\frac{n^4}{\varepsilon^2}\log^9\frac{n}{\varepsilon\delta} + n^4\log^8\frac{n}{\delta}\log R\right) = O^*(n^4)$$

oracle calls.

The number of arithmetic operations is $O^*(n^6)$, 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. In the next section, we describe the volume algorithm. For the analysis (Sec. 2.4 gives an outline), we will need some tools about logconcavity and probability (Sec. 3). In the description of the volume algorithm, we will assume that $B \subseteq K \subseteq O^*(\sqrt{n})B$. In Sec. 5, we show how to achieve this by an algorithm that "rounds" a given convex body.

2 The volume algorithm

In this section, we describe the main volume algorithm. We will assume that the convex body of interest, $K \subseteq \mathbb{R}^n$, contains the unit ball B and is contained in the ball DB, where $D = O(\sqrt{n}\ln(1/\varepsilon))$. If this is not true for the given K, it can be achieved by a pre-processing step (finding and applying a suitable affine transformation) which is described in Section 5. To avoid some trivial difficulties, we assume that $n \geq 16$.

The main part of the algorithm consists of a modification of K, called the "pencil" construction, described in Sec. 2.2, followed by a multi-phase Monte-Carlo estimation. The algorithm uses a subroutine for convex body sampling as a black box, described next.

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 $f(x)$ is proportional to a given exponential function $e^{-a^T x}$, i.e.,

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

Let the corresponding measure be μ_f. 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 $X \in K$ drawn from some measure σ, a bound M on the L_2 norm of σ w.r.t. μ_f, and an accuracy parameter $\delta > 0$;
- *Output:* a point $Y \in K$ whose measure has variation distance at most δ from μ_f.

A sampler we can use is given in [18], using the hit-and-run algorithm. We make the following assumptions about the input:

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

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

(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

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

A distribution that satisfies (A1) and (A2) is said to be (r, R)-*rounded*. If $r = \Omega(1)$ and $R = O^*(1)$, then we say that f is *well-rounded*.

Under these assumptions, the main result of [18] says that the total variation distance of the output distribution from μ_f is less than δ. Furthermore, the number of calls on the membership oracle is

$$O\left(n^2\frac{R^2}{r^2}\ln^5\frac{Mn}{\delta^2}\right).$$

In the analysis, we will show that (A1) is satisfied with $r \geq 1/10$, (A2) is satisfied with $R = O(\sqrt{n}\ln\frac{1}{\varepsilon})$ and (A3) with $M \leq 8$. Thus, the complexity is $O(n^3\ln^7\frac{n}{\varepsilon\delta})$ per random point.

For the special case when $a = 0$, i.e., f is the uniform distribution, we can simplify (A1) to the condition that K contains a ball of radius r. In this case, the sampler is a bit more efficient — the number of calls to the membership oracle is

$$O\left(n^2\frac{R^2}{r^2}\ln^3\frac{M}{\delta}\right).$$

2.2 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

$$C = \{x \in \mathbb{R}^{n+1} \ : \ x_0 \geq 0, \sum_{i=1}^n x_i^2 \leq x_0^2\}$$

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 tip is at the origin. Note that by the definition of D, the part of K' in the halfspace $x_0 \geq D$ is inside C and so it is a cylinder over K. Also, since K contains a unit ball, the part of K' in the halfspace $x_0 \leq 1$ is a cone C_B over the unit ball. See Fig. 1 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 $O(1/\varepsilon^2)$ 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 contained in a ball of radius $2D$.

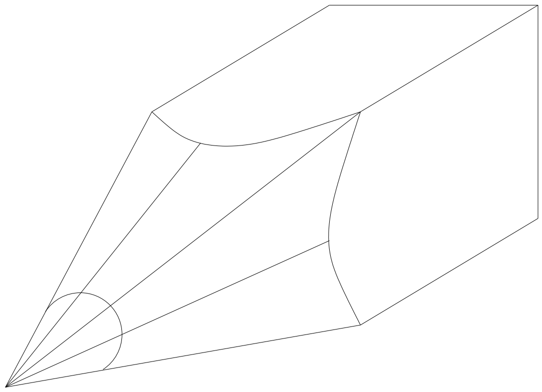

Figure 1: The pencil construction when K is a square.

2.3 The multi-phase Monte-Carlo

Now we describe the "morphing" part of the algorithm. 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 $a \le \varepsilon^2/D$, an easy computation shows that

$$(1-\varepsilon)\mathrm{vol}(K') \le Z(a) \le \mathrm{vol}(K'),$$

so it suffices to compute $Z(a)$ for such an a. On the other hand, for $a \ge 2n$ the value of $Z(a)$ is essentially the same as the integral over the whole cone, which is easy to compute:

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

(where $\pi_n = \mathrm{vol}(B)$) and

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

by standard computation.

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

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

The algorithm will estimate $Z(a_m)$ by estimating the ratios

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

We will estimate the R_i's using sampling. Let μ_i be the probability distribution over K' with density proportional to $e^{-a_i x_0}$, i.e.,

$$\frac{d\mu_i(x)}{dx} = \frac{e^{-a_i x_0}}{Z(a_i)}.$$

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:

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

So, to estimate the ratio R_i, we draw random samples $X^1, \ldots, X^k$ from μ_i, and compute the average of the corresponding Y's.

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

Sample points that are (approximately) 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 $X = (X_0, X_0 Y_1, \ldots, X_0 Y_n) \notin K'$, try again; else, return X.

In order to get appropriate sample points from μ_i ($i > 0$) efficiently (i.e., satisfy assumptions (A1) and (A2) of the sampler), we have to make a simple affine transformation, namely a scaling along the x_0 axis. Let $\gamma_i = \max(1, a_i/\sqrt{n})$, and

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

This will ensure that the distributions we sample are well-rounded. 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. For calling the oracle, the roundness parameters r, R and the warm start measure M will always be bounded by the same values, and so we do not mention them below. The 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 $m = 2\lceil \sqrt{n} \ln \frac{n}{\varepsilon} \rceil$, $\quad k = \frac{512}{\varepsilon^2} \sqrt{n} \ln \frac{n}{\varepsilon}$, $\quad \delta = \varepsilon^2 n^{-10}$ and

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

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

— Run the sampler k times for convex body $T_i K$, with vector $(a_i/\gamma_i, 0, \ldots, 0)$ (i.e., exponential function $e^{-a_i x_0/\gamma_i}$), error parameter δ, and (for $i > 0$) starting points $T_i X_{i-1}^1, \ldots T_i X_{i-1}^k$. 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

$$Z = n! \pi_n (2n)^{-(n+1)} W_1 \ldots W_m$$

as the estimate of the volume of K'.

2.4 Outline of analysis

The analysis of the algorithm will verify the following claims. The first asserts that estimate computed by the algorithm is accurate, while the second and third address the conditions required by the sampler.

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) and the dependence between samples is small enough that the output is accurate with large probability (Lemma 4.2).

2. Random samples from one phase provide a good start for the next phase (Lemma 4.4), i.e., the warm start measure M is bounded and (A3) is satisfied.

3. The convex body $T_i K$ and exponential function $e^{-a_i x_0/\gamma_i}$ satisfy (A1) and (A2) (Lemmas 4.5, 4.6), the roundness assumptions required by the convex body sampler.

4. The overall complexity is

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

Together, these claims yield Theorem 1.1. Their proofs need some technical tools, collected in the next section (they might be useful elsewhere).

Below, we are going to assume that $n \geq 15$ and $\varepsilon > (3/4)^n$, which only excludes uninteresting cases but makes the formulas simpler.

3 Analysis tools

3.1 Logconcavity

A nonnegative function $f : \mathbb{R}^n \to \mathbb{R}_+$ is said to be *logconcave* if for any $x, y \in \mathbb{R}^n$ and any $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}.$$

The following theorem proved by Dinghas [4], Leindler [11] and Prékopa [19, 20] summarizes fundamental properties of logconcave functions.

Theorem 3.1 *All marginals as well as the distribution function of a logconcave function are logconcave. The convolution of two logconcave functions is logconcave.*

The next lemma will play a key role in the analysis.

Lemma 3.2 *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. By Theorem 3.1, its marginal in t is a logconcave function of t. But this marginal is just

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

□

The next lemma is from [16].

Lemma 3.3 *Let X be a random variable with a logconcave distribution, such that $\mathsf{E}(X) = \bar{X}$ and $\mathsf{E}(|X - \bar{X}|^2) = \sigma^2$. Then,*

$$\mathsf{P}\left(|X - \bar{X}| > t\sigma\right) \leq e^{-t}.$$

The following lemma is along the lines of Theorem 4.1 in [8] and its proof is the same.

Lemma 3.4 *Let K be a convex body in $\mathbb{R}^n$ whose centroid is the origin. For a unit vector v, let*

$$-a = \min_{x \in K} v^T x, \quad b = \max_{x \in K} v^T x \quad \textit{and} \quad \sigma^2 = \mathsf{E}_K((v^T x)^2).$$

$$\textit{Then,} \quad \sigma\sqrt{\frac{n+2}{n}} \leq a, b \leq \sigma\sqrt{n(n+2)}.$$

3.2 Probability

For two random variables X, Y, we will use the following measure of their independence:

$$\mu(X,Y) = \sup_{A,B} \left| \mathsf{P}(X \in A,\ Y \in B) - \mathsf{P}(X \in A)\mathsf{P}(Y \in B) \right|,$$

where A and B range over measurable subsets of the ranges of X and Y. By the identity

$$\left|\mathsf{P}(X \in A,\ Y \in B) - \mathsf{P}(X \in A)\mathsf{P}(Y \in B)\right| = \left|\mathsf{P}(X \in \overline{A},\ Y \in B) - \mathsf{P}(X \in \overline{A})\mathsf{P}(Y \in B)\right|, \quad (3)$$

it suffices to consider sets A, B with $\mathsf{P}(X \in A) \geq 1/2$ and $\mathsf{P}(Y \in B) \geq 1/2$.

The following are some basic properties of μ.

Lemma 3.5 *If f and g are two measurable functions, then*

$$\mu(f(X), g(Y)) \leq \mu(X, Y).$$

Proof. Trivial from the equation

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

□

The next lemma is a variation of a lemma in [9].

Lemma 3.6 *Let X, Y be random variables such that $|0 \leq X \leq a$ and $0 \leq Y \leq b$. Then*

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

Proof. The LHS can be written as

$$\left| \int_0^a \int_0^b \mathsf{P}(X \geq x, Y \geq y)\, dx dy \right|$$

which is clearly at most μab. □

Lemma 3.7 *Let X, Y, X', Y' be random variables, and assume that the pair (X, Y) is independent from the pair (X', Y'). Then*

$$\mu((X, X'), (Y, Y')) \leq \mu(X, Y) + \mu(X', Y').$$

Proof. Set $\mu = \mu(X, Y)$ and $\mu' = \mu(X', Y')$. Let R, R', S, S' be the ranges 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

$$\left|\mathsf{P}((X, X') \in A,\ (Y, Y') \in B) - \mathsf{P}((X, X') \in A)\mathsf{P}((Y, Y') \in B)\right| \leq \mu + \mu'. \quad (4)$$

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

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

(here we use that (X,Y) is independent of (X',Y')). We can write the left hand side of (4) as

$$\begin{aligned}&\mathsf{E}(h(X,Y))-\mathsf{E}(f(X))\mathsf{E}(g(Y))\\ &=\Big[\mathsf{E}(h(X,Y)-f(X)g(Y))\Big]+\Big[\mathsf{E}(f(X)g(Y))-\mathsf{E}(f(X))\mathsf{E}(g(Y))\Big]\end{aligned}\tag{5}$$

By assumption,

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

for every r and s, and hence the first term on the right hand side in (5) is at most μ' in absolute value. The second term is at most μ by Lemma 3.6. This proves (4). □

Lemma 3.8 *Let $X_0,X_1,\ldots,$ be a Markov chain. Then*

$$\mu(X_i,X_{i+1})=\mu((X_1,\ldots,X_i),X_{i+1}).$$

Proof. The inequality $\le$ is trivial (e.g. by Lemma 3.5). To show the converse, set $\mu=\mu(X_i,X_{i+1})$. 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

$$\big|\mathsf{P}((X_0,\ldots,X_i)\in A,\ X_{i+1}\in B)-\mathsf{P}((X_0,\ldots,X_i)\in A)\mathsf{P}(X_{i+1}\in B)\big|\ \le\ \mu.\tag{6}$$

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 $r\in S_i$,

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

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 (6) follows from Lemma 3.6 again. □

We close this section with another simple fact from probability.

Lemma 3.9 *Let $X\ge 0$ be a random variable, $a>0$, and $X'=\min(X,a)$. Then*

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

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

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

which implies the assertion. □

4 Analysis of volume algorithm

The analysis is divided into three sections. The first two sections establish that the answer found by the algorithm is accurate — Sec. 4.1 shows that the variance of each random variable computed is small, using Lemma 3.2 about logconcavity; Sec. 4.2 shows how to handle the (slight) dependence among the random variables. The latter section is somewhat technical and uses the properties of μ-independence developed in Sec. 3. Together, they imply that the number of iterations and samples used by the algorithm are sufficient.

The last section is devoted to proving properties needed by the convex body sampler, thus ensuring that the sampling is efficient. Lemma 4.4 shows that samples from one phase have a bounded warm start measure for the next distribution (i.e., assumption (A3) is satisfied) and Lemmas 4.5, 4.6 show that the distribution being sampled remains well-rounded (i.e., assumptions (A1) and (A2) are satisfied) throughout the course of the algorithm.

4.1 Variance

We begin by bounding the variance of the random variables used to estimate the ratios R_i defined by (1).

Lemma 4.1 *Let X be a random sample from $d\mu_i$ and*

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

Then

$$\frac{\mathsf{E}(Y^2)}{\mathsf{E}(Y)^2} \le \left(\frac{a_{i+1}^2}{a_i(2a_{i+1}-a_i)}\right)^{n+1} < 8.$$

Proof. We have

$$\mathsf{E}(Y) = \frac{\int_{K'} e^{-a_{i+1}x_0}\,dx}{\int_{K'} e^{-a_i x_0}\,dx}$$

and

$$\mathsf{E}(Y^2) = \frac{\int_{K'} e^{-(2a_{i+1}-a_i)x_0}\,dx}{\int_{K'} e^{-a_i x_0}\,dx}.$$

By Lemma 3.2, the function $a^{n+1}\int_{K'} e^{-ax_0}\,dx$ is logconcave and so

$$\int_{K'} e^{-a_i x_0}\,dx \int_{K'} e^{-(2a_{i+1}-a_i)x_0}\,dx \le \left(\frac{a_{i+1}^2}{a_i(2a_{i+1}-a_i)}\right)^{n+1}\left(\int_{K'} e^{-a_{i+1}x_0}\,dx\right)^2.$$

Thus, since $a_{i+1} = a_i(1-1/\sqrt{n})$,

$$\frac{E(Y^2)}{E(Y)^2} \le \left(\frac{a_{i+1}^2}{a_i(2a_{i+1}-a_i)}\right)^{n+1} = \left(1+\frac{1}{n-2\sqrt{n}}\right)^{n+1},$$

and the lemma follows. $\square$

Remark: If we have $a_{i+1} = a_i(1-1/t)$, then

$$\frac{\mathsf{E}(Y^2)}{\mathsf{E}(Y)^2} \le \left(1+\frac{1}{t^2-2t}\right)^{n+1}.$$

4.2 Divine intervention

In this section, our goal is to prove that the volume estimate is accurate. Note that in the lemma below, the probability of success can be boosted from $3/4$ to $1-\delta$ for any $\delta > 0$, by using the standard trick of repeating the algorithm $O(\log(1/\delta))$ times and taking the median of the outputs.

Lemma 4.2 *With probability at least* $3/4$*,*

$$\left(1-\frac{\varepsilon}{2}\right) R_1 \ldots R_m \le W_1 \ldots W_m \le \left(1+\frac{\varepsilon}{2}\right) R_1 \ldots R_m.$$

Proof. 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.

To analyze the algorithm we use the "divine intervention" (a.k.a. coupling) method. The distribution of the random point X_i^j is *approximately* μ_i. We construct modified random variables $\overline{X}_i^j$ ($i = 0, \ldots, m$, $j = 1, \ldots, k$) 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+1} 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

$$\mathsf{P}(\overline{X}_{i+1}^j = X_{i+1}^j) \ge 1-\delta. \tag{7}$$

It follows by induction that $X_i^j = \overline{X}_i^j$ with probability at least $1-j\delta$, and hence

$$\mathsf{P}(X_i^j = \overline{X}_i^j \text{ for all } j) \ge 1-km\delta. \tag{8}$$

[It would be nice to use "divine intervention" to make the X_i^j in one of the threads 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.]

Let

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

Since the random variables $\overline{X}_i^j$ have the "right" distribution μ_i, we have $\mathsf{E}(Y_i^j) = R_i$. Let

$$\overline{W}_i = \frac{1}{k}\sum_{j=1}^{k} Y_i^j.$$

Then for every j, $\mathsf{E}(\overline{W}_i) = \mathsf{E}(Y_i^j) = R_i$, and using Lemma 4.1, $\mathsf{E}((Y_i^j)^2) \le 8\mathsf{E}(Y_i^j)^2$. Thus,

$$\mathsf{E}(\overline{W}_i^2) = \frac{1}{k^2}\left(\sum_{j=1}^{k} \mathsf{E}((Y_i^j)^2) + k(k-1)R_i^2\right) \le \left(1+\frac{7}{k}\right) R_i^2. \tag{9}$$

Consider the product $\overline{Z} = \overline{W}_1\overline{W}_2\dots\overline{W}_m$. Now *if* we had independence between successive phases, then we would have

$$\mathsf{E}(\prod_{i=1}^{m}\overline{W}_i) = \prod_{i=1}^{m} R_i \quad \text{and} \quad \mathsf{E}(\prod_{i=1}^{m}\overline{W}_i^2) \le \left(1+\frac{7}{k}\right)^m \prod_{i=1}^{m} R_i^2.$$

This would imply

$$\text{var}(\overline{Z}) \le (1+\frac{7}{k})^m - 1 \le \frac{\varepsilon^2}{32}\prod_{i=1}^{m} R_i^2,$$

and we could easily show that with probability at least $4/5$ (say), $\overline{Z}$ 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 $3/4$, $\overline{Z}$ 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 not independent. The next lemma shows that they are approximately independent; its proof is deferred to the end of the section.

Lemma 4.3 (a) *For every phase $0 \le i < m$ and every thread $1 \le j \le k$, 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 $0 \le i < m$ and every thread $1 \le j \le k$, the random variables $(X_0^j, \dots, X_i^j)$ and X_{i+1}^j are (3δ)-independent.*

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

We continue with the proof of Lemma 4.2. Given the approximate independence, we would like to apply Lemma 3.6 to bound the expectation. However, the variables $\overline{W}_1 \dots \overline{W}_i$ and $\overline{W}_{i+1}$ are not bounded and so we cannot apply Lemma 3.6 directly. So we define another set of random variables. Let

$$\alpha = \frac{\varepsilon^{\frac{1}{2}}}{8(m\mu)^{\frac{1}{4}}}$$

where $\mu = 3k\delta$. Note that α is much larger than 1. For $i = 1, \dots, m$, let

$$V_i = \min\{\overline{W}_i, \alpha\mathsf{E}(\overline{W}_i)\}.$$

By our choice of δ, $V_i = \overline{W}_i$ with high probability.

Clearly $\mathsf{E}(V_i) \le \mathsf{E}(\overline{W}_i)$. On the other hand, using Lemma 3.9 and equation (9), we have

$$\mathsf{E}(V_i) \ge \mathsf{E}(\overline{W}_i) - \frac{\mathsf{E}(\overline{W}_i^2)}{4\alpha\mathsf{E}(\overline{W}_i)} \ge (1 - \frac{1}{4\alpha}\left(1+\frac{7}{k}\right)^m)\mathsf{E}(\overline{W}_i) \ge (1-\frac{1}{2\alpha})\mathsf{E}(\overline{W}_i). \tag{10}$$

Define $U_0 = 1$ and recursively

$$U_{i+1} = \min\{U_iV_{i+1}, \alpha\mathsf{E}(V_1)\dots\mathsf{E}(V_{i+1})\}.$$

We will show that

$$\left(1 - \frac{(i-1)}{\alpha}\right)\mathsf{E}(V_1)\dots\mathsf{E}(V_i) \le \mathsf{E}(U_i) \le \left(1 + 2\mu\alpha^2 i\right)\mathsf{E}(V_1)\dots\mathsf{E}(V_i). \tag{11}$$

By Lemma 3.5, the random variables U_i and V_{i+1} are μ-independent, so it follows by Lemma 3.6 that

$$|\mathsf{E}(U_i V_{i+1}) - \mathsf{E}(U_i)\mathsf{E}(V_{i+1})| \le \mu\alpha \mathsf{E}(V_1)\ldots\mathsf{E}(V_i)\alpha\mathsf{E}(\overline{W}_{i+1}) \le 2\mu\alpha^2\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1}). \quad (12)$$

Now the upper bound in (11) is easy by induction:

$$\begin{aligned}\mathsf{E}(U_{i+1}) \le \mathsf{E}(U_iV_{i+1}) &\le \mathsf{E}(U_i)\mathsf{E}(V_{i+1}) + 2\mu\alpha^2\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})\\ &\le \big(1+2\mu\alpha^2(i+1)\big)\,\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1}) \qquad \text{(induction hypothesis)}.\end{aligned}$$

A similar argument shows that

$$\mathsf{E}(U_i^2) \le \big(1+2\mu\alpha^4 i\big)\,\mathsf{E}(V_1^2)\ldots\mathsf{E}(V_i^2) \quad (13)$$

$$\text{and} \quad \mathsf{E}(U_i^2V_{i+1}^2) \le \big(1+2\mu\alpha^4 i\big)\,\mathsf{E}(V_1^2)\ldots\mathsf{E}(V_{i+1}^2). \quad (14)$$

For the lower bound in (11), using Lemma 3.9 and inequality (14), we get

$$\begin{aligned}\mathsf{E}(U_{i+1}) &\ge \mathsf{E}(U_iV_{i+1}) - \frac{\mathsf{E}(U_i^2V_{i+1}^2)}{4\alpha\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})}\\ &\ge \mathsf{E}(U_iV_{i+1}) - \big(1+2\mu\alpha^4 i\big)\,\frac{\mathsf{E}(V_1^2)\ldots\mathsf{E}(V_{i+1}^2)}{4\alpha\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})}.\end{aligned}$$

Here, using (10),

$$\begin{aligned}\mathsf{E}(V_i^2) \le \mathsf{E}(\overline{W}_i^2) &\le \left(1+\frac{7}{k}\right)\mathsf{E}(\overline{W}_i)^2\\ &\le \left(1+\frac{7}{k}\right)\frac{1}{1-2\sqrt{n\mu}}\mathsf{E}(V_i)^2\\ &< \left(1+\frac{8}{k}\right)\mathsf{E}(V_i)^2. \end{aligned} \quad (15)$$

Hence,

$$\begin{aligned}\mathsf{E}(U_{i+1}) &\ge \mathsf{E}(U_iV_{i+1}) - \frac{1}{4\alpha}\big(1+2\mu\alpha^4 i\big)\left(1+\frac{8}{k}\right)^i\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})\\ &\ge \mathsf{E}(U_iV_{i+1}) - \frac{1+2\mu\alpha^4 i}{2\alpha}\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})\\ &\ge \mathsf{E}(U_i)\mathsf{E}(V_{i+1}) - \frac{1}{\alpha}\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1})\end{aligned}$$

(we used (12) in the last step). Hence, by induction

$$\mathsf{E}(U_{i+1}) \ge \mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1}) - \frac{i}{\alpha}\mathsf{E}(V_1)\ldots\mathsf{E}(V_{i+1}).$$

This proves (11). Thus,

$$\mathsf{E}(U_m) \le \left(1+\frac{\varepsilon}{4}\right)\mathsf{E}(V_1)\ldots\mathsf{E}(V_m) \le \left(1+\frac{\varepsilon}{4}\right)\mathsf{E}(\overline{W}_1)\ldots\mathsf{E}(\overline{W}_m).$$

Similarly,

$$\mathsf{E}(U_m) \geq \left(1 - \frac{\varepsilon}{4}\right) \mathsf{E}(\overline{W}_1) \dots \mathsf{E}(\overline{W}_m).$$

By (13) and (15),

$$\mathsf{E}(U_m^2) \leq \left(1 + \frac{\varepsilon^2}{64}\right) \mathsf{E}(U_m)^2,$$

and hence

$$|U_m - \mathsf{E}(U_m)| \leq \frac{\varepsilon}{2} \mathsf{E}(\overline{W}_1) \dots \mathsf{E}(\overline{W}_m)$$

with probability at least .9. Furthermore, using Markov's inequality,

$$\mathsf{P}(U_{i+1} \neq U_i V_{i+1}) = \mathsf{P}\left(U_i V_{i+1} > \alpha \mathsf{E}(V_1) \dots \mathsf{E}(V_{i+1})\right) \leq \frac{2}{\alpha}$$

and similarly

$$\mathsf{P}(V_i \neq \overline{W}_i) \leq \frac{1}{\alpha}.$$

So with probability at least $1 - \frac{3k}{\alpha}$, we have $U_m = \overline{W}_1 \dots \overline{W}_m$. Also, by (8), we have $\overline{W}_1 \dots \overline{W}_m = W_1 \dots W_m$ with probability at least $1 - km\delta$.

Note that $\mathsf{E}(\overline{W}_1) \dots \mathsf{E}(\overline{W}_m) = R_1 \dots R_m$. Hence with probability at least $3/4$

$$|W_1 \dots W_m - R_1 \dots R_m| \leq \frac{\varepsilon}{2} R_1 \dots R_m.$$

This completes the proof of Lemma 4.2. □

Proof. (of Lemma 4.3.)

(a) Let $A, B \subseteq K$; we claim that

$$\left|\mathsf{P}(X_i^j \in A, X_{i+1}^j \in B) - \mathsf{P}(X_i^j \in A)\mathsf{P}(X_{i+1}^j \in B)\right| \leq \delta. \tag{16}$$

By the remark after (3), 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 so $\|\mu_i'/\mu_{i+1}\| \leq 4\|\mu_i/\mu_{i+1}\|$. 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 (16) holds. The second assertion is immediate, since putting a bar above the variables changes the probabilities in the condition by at most δ (cf. (7)).

(b) Follows from Lemma 3.8.

(c) Follows from Lemmas 3.7 and 3.5. □

4.3 Sampling assumptions

The next lemma shows that samples from one phase provide a good start for the next phase. This means that assumption (A3) is satisfied in every phase of the algorithm.

Lemma 4.4 *The L_2-norm of μ_i with respect to μ_{i+1} is at most* 8.

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

$$\mathsf{E}\left[\frac{d\mu_i(X)}{d\mu_{i+1}(X)}\right] \le 8.$$

Indeed, using Lemma 3.2

$$\begin{aligned}
\mathsf{E}\left[\frac{d\mu_i(X)}{d\mu_{i+1}(X)}\right] &= \frac{\int_{K'} e^{(a_{i+1}-a_i)x_0} e^{-a_i x_0}\,dx}{\int_{K'} e^{-a_i x_0}\,dx} \frac{\int_{K'} e^{-a_{i+1}x_0}\,dx}{\int_{K'} e^{-a_i x_0}\,dx} \\
&= \frac{Z(2a_i - a_{i+1})Z(a_{i+1})}{Z(a_i)Z(a_i)} \\
&\le \left(\frac{(2a_i)^2}{4a_{i+1}(2a_i - a_{i+1})}\right)^{n+1} \\
&= \left(\frac{1}{(1-\frac{1}{\sqrt{n}})(1+\frac{1}{\sqrt{n}})}\right)^{n+1} < 8.
\end{aligned}$$

□

Finally, we show that (A1) and (A2) are maintained in the algorithm. For $s \ge 0$, let

$$K_s = \{x \in K' :\ x_0 \le s\}.$$

These sets are exactly the level sets of the functions $f_i = e^{-a_i x_0}$.

Lemma 4.5 *Let $c = \mu_i(K_s)$. Then $T_i K_s$ 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 $T_i K'$ contains a cone with height $\gamma_i s \ge s$ over this ball. If we show that $s \ge c/4$, then it follows by simple geometry that K_s contains a ball with radius $c/4(1+\sqrt{2}) > c/10$.

We may assume that $s < 1/4$. Let $F(t)$ denote the density function of X_0. This function is proportional to $t^n e^{-a_i t}$ 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 $\mathsf{E}(|T_i X|^2) \le 5D^2$.*

Proof. Let $X = (X_0, \ldots, X_n)$ be a random point from μ_i, and let $Y = (Y_0, \ldots, Y_n) = T_i X$. 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 contained in DB. This proves the lemma. □

5 Rounding the body

In the volume algorithm, we assumed that K contains the unit ball B and is contained in the ball DB, where $D = O(\sqrt{n}\ln(1/\varepsilon))$. Here we describe an algorithm to achieve this, which can be used as a pre-processing step if necessary. In [9], a similar procedure for rounding had to be interlaced with volume computation for efficiency. Here too one can interlace, but it is conceptually easier and just as efficient to view the rounding as a pre-processing step. In fact, the rounding will only require uniform sampling. The rounding algorithm actually finds an affine tranformation that puts K in near-isotropic position (defined below). As will be clear shortly, this is a stronger property than being well-rounded.

A convex body K is said to be in *isotropic* position, if its center of gravity is the origin and for every vector $v \in \mathbb{R}^n$,

$$\int_K (v \cdot x)^2 \, dx = |v|^2.$$

In terms of the associated random variable X, this means that

$$\mathsf{E}_K(X) = 0 \quad \text{and} \quad \mathsf{E}_K(XX^T) = I.$$

We say that K is *near-isotropic up to a factor* t, if

$$\frac{1}{t} \le \int_K (u^T(x - \bar{x}))^2 \, dx \le t$$

for every unit vector u. Recall that a convex set K is (r, R)-rounded if (i) it contains a ball of radius r and (ii) $\mathsf{E}_K(|x - \bar{x}|^2) \le R^2$.

A body in isotropic position has many nice roundness properties. For example, it follows from the definition, that

$$\mathsf{E}_K(|X|^2) = \sum_{i=1}^n \mathsf{E}_K(X_i^2) = n$$

and from Lemma 3.4 it contains a ball of radius 1 and so it is $(1, \sqrt{n})$-rounded. Further, by Lemma 3.3,

$$\mathsf{P}\left(|X| > s\sqrt{\mathsf{E}_K(|X|^2)}\right) \le e^{-s}. \tag{17}$$

For a body in t-isotropic position, $\mathsf{E}_K(|X - \bar{X}|^2) \le tn$ and so, all but an ε fraction is contained in a ball of radius $\sqrt{tn}\ln(1/\varepsilon)$. Thus, if we first transformed K into near-isotropic position for some constant t and then replace it by its intersection with a ball of radius D for $D > \sqrt{tn}(\ln(1/\varepsilon) + 2)$, we satisfy the assumptions of the algorithm and lose less than an $\frac{\varepsilon}{e^2}$ fraction of its volume. The rest of this section is devoted to an algorithm that brings a given convex body into near-isotropic position.

It is well-known that there is an affine transformation that will put any convex body in isotropic position. In fact, it can be estimated as follows: we generate $O(n \log^2 n)$ approximately uniformly distributed random points in K, and compute the transformation that brings these points (more exactly, the uniform measure on this finite set) into isotropic position. More precisely, if $y_1, \ldots, y_l$ are uniform random points from K, we compute:

$$\bar{y} = \frac{1}{l}\sum_{i=1}^{l} y_i \quad Y = \frac{1}{l}\sum_{i=1}^{l}(y_i - \bar{y})(y_i - \bar{y})^T \quad \text{and } T = Y^{-\frac{1}{2}}. \tag{18}$$

We make use of the following theorem due to Rudelson [21].

Theorem 5.1 *[21] Let K be a convex body in $\mathbb{R}^n$ in isotropic position and $\eta > 0$. Let $y_1, \ldots, y_l$ be independent random points distributed uniformly in K, with*

$$l \geq c\frac{np}{\eta^2} \log \frac{n}{\eta^2} \max\{p, \log n\}.$$

where c is a fixed absolute constant and p is any positive integer. Then,

$$\mathsf{E}\left(\left\|\frac{1}{l}\sum_{i=1}^{l} y_i y_i^T - I\right\|^p\right) \leq \eta^p.$$

A convenient way to use this is the following corollary (formulated in [3]).

Corollary 5.2 *Let K be a convex set and $\bar{y}$ and Y be defined as in (18). There is an absolute constant c such that for any integer $p \geq 1$, and $N \geq cpn \log n \max\{p, \log n\}$, with probability at least $1 - \frac{1}{2^{p-1}}$, the set*

$$K_1 = \{x : \; Y^{\frac{1}{2}}x + \bar{y} \in K\}$$

satisfies

$$\frac{1}{2}||v||^2 \leq \mathsf{E}_{K_1}((v^T x)^2) \leq \frac{3}{2}||v||^2.$$

Thus uniform samples are enough to calculate a near-isotropic transformation. But how do we get uniform samples?! After all, it is in order to make the sampling efficient that we are rounding in the first place. To get around this, we will alternate between sampling and rounding steps. We will also use the pencil construction.

We assume that K contains the unit ball B and is contained in $R_0 B$. As in Section 2.2, let C be the cone in $\mathbb{R}^{n+1}$ defined by

$$C = \{x \in \mathbb{R}^{n+1} \; : \; x_0 \geq 0, \sum_{i=1}^{n} x_i^2 \leq x_0^2\}.$$

Then $K' \in \mathbb{R}^{n+1}$ is a pencil defined as

$$K' = \Big([0, 2R_0] \times K\Big) \cap C.$$

Also define for $i = 1, \ldots, \lceil n \log 2R_0 \rceil$,

$$K_i = K' \cap \{x \,|\, x_0 \leq 2^{\frac{i}{n}}\}.$$

Since K contains a unit ball centered at the origin, K_0 is just a rotational cone, and it is easy to generate a random point from it. The output of the algorithm below is a linear transformation that puts K in near-isotropic position.

In every phase, the algorithm scales along the x_0 axis using a small number of samples. Once every n phases, it applies a linear transformation orthogonal to the x_0 axis. (The goal of these phases is to keep the body well rounded so that sampling is efficient.) The cumulative transformation it computes is of the form

$$S_i = \begin{pmatrix} t_i & 0 \\ 0 & T_i \end{pmatrix},$$

where t_i is a positive number and T_i is an $n \times n$ matrix.

Rounding algorithm:

R1. Set $m = \lceil n \log 2R_0 \rceil$ and $\delta = \varepsilon^{10} n^{-10}$, and initialize $t_0 = n$ and $T_0 = \sqrt{n} I_n$. Let X_0 is a random point from K_0.

R2. For $i = 0, \ldots, m-1$, do the following.

If i is *not* a multiple of n,

— Run the uniform sampler $k = 2c \ln n$ times for convex body $S_i K_i$, with error parameter δ, and initial starting point $S_i X_{i-1}$. Use each point obtained as the next starting point. Apply S_i^{-1} to the points obtained to get points $Y^1, \ldots, Y^k$ in K_i. Set $X_i = Y^k$.

— Using the sample points, compute

$$\bar{Y}_0 = \frac{1}{k} \sum_{j=1}^{k} (Y^j)_0 \quad \text{and} \quad Z = \frac{1}{k} \sum_{j=1}^{k} (Y^j)_0^2 - \bar{Y}_0^2.$$

— For $p, q \in \{0, 1, \ldots, n\}$

$$t_{i+1} = \frac{1}{\sqrt{Z}},$$

and let $T_{i+1} = T_i$.

Else (if i is a multiple of n),

— Run the uniform sampler $l = 2cn \log^3 n$ times for convex body $S_i K_i$, with error parameter δ, and initial starting point $S_i X_{i-1}$. Use each point obtained as the next starting point. Apply S_i^{-1} to the points obtained to get points $Y^1, \ldots, Y^l$ in K_i. Set $X_i = Y^l$. Let $\hat{Y}^j$ be obtained from Y^j by dropping the coordinate with index 0.

— Using these points, compute

$$\bar{Y} = \frac{1}{l} \sum_{j=1}^{l} \hat{Y}^j \quad \text{and} \quad A = \frac{1}{l} \sum_{j=1}^{l} \hat{Y}^j (\hat{Y}^j)^T - \bar{Y} \bar{Y}^T.$$

— Compute $T_{i+1} = A^{-\frac{1}{2}}$, and let $t_{i+1} = t_i$.

R3. Run the uniform sampler $2l$ times on $S_m K'$ and apply S_m^{-1} to the resulting points. Let the subset with $x_0 \geq R$ be $Y^1, \ldots, Y^r$. Let $\hat{Y}^j$ be obtained from Y^j by dropping the coordinate with index 0. Compute

$$\bar{Y} = \frac{1}{r} \sum_{j=1}^{r} \hat{Y}^j \quad \text{and} \quad A_{pq} = \frac{1}{r} \sum_{j=1}^{r} \hat{Y}^j \hat{Y}^j - \bar{Y} \bar{Y}^T.$$

Return the matrix $T = A^{-1/2}$.

Theorem 5.3 *The set TK is in 2-isotropic position with high probability. Further, the number of* *calls to the membership oracle is*

$$O(n^4 \log^8 n \log R) = O^*(n^4).$$

Theorem 5.3 will follow from the next lemma which shows that the assumptions of the sampler are satisfied so that each sample requires only $O(n^3 \log^5 n)$ oracle calls. The total number of samples is $O(n \log^3 n \log R_0)$.

Lemma 5.4 *With probability at least $1 - \eta$, for $i = 1, \ldots, m$, the sets $S_i K_{i-1}$ and $S_i K_i$ are $(\frac{1}{4}, 6\sqrt{n(\ln^2 2n + 2)})$-rounded.*

Proof. First note that each transformation S_i can be viewed as a composition of two transformations — a scaling along the x_0 axis and a linear transformation orthogonal to the x_0 axis. As a result, the distribution of $S_i K_i$ along the x_0-axis is just a scaled version of the distribution of K_i. In particular, it is monotone nondecreasing. Also, since K' is a pencil, for any $t' \geq t$ such that the cross-section $K_i \cap \{x : x_0 = t'\}$ is nonempty, we have

$$(t' - t, 0, \ldots, 0) + K_i \cap \{x : x_0 = t_1\} \subseteq K_i \cap \{x : x_0 = t'\}.$$

The proof will proceed by induction on i. The sets $S_1 K_1$ and $S_1 K_0$ are just scaled rotational cones. Suppose $S_{i-1} K_{i-1}$ is $(1/4, 6\sqrt{n(\ln^2 n + 2)})$-rounded. We will first prove that $S_i K_{i-1}$ and $S_i K_i$ contain balls of radius $1/4$.

If i is a multiple of n, the algorithm obtains $2l$ nearly uniform samples from K_{i-1} to estimate the inertia matrix A, and so by Corollary 5.2, with high probability for every vector v orthogonal to the x_0 axis,

$$\frac{1}{2} \leq \mathsf{E}_{S_i K_{i-1}} \left((v^T (x - \bar{x}))^2 \right) \leq \frac{3}{2}. \tag{19}$$

Hence, by Lemma 3.4, the projection of $S_i K_{i-1}$ orthogonal to the x_0 axis contains an n-dimensional ball of radius $1/\sqrt{2}$ centered at the the projection of z_i, the centroid of $S_i K_{i-1}$. Let $u = \max\{x_0 : x \in S_i K_i\}$. In particular, the cross-section of $S_i K_{i-1}$ at $x_0 = u$ contains such a ball. Now for the x_0 axis itself, we get

$$\frac{1}{2} \leq \mathsf{E}_{S_i K_{i-1}} \left((x - \bar{x})_0^2 \right) \leq \frac{3}{2}. \tag{20}$$

Hence, by Lemma 3.4, $|u - (z_i)_0| \geq 1/\sqrt{2}$ and so $S_i K_{i-1}$ contains a cone of height $1/\sqrt{2}$. A cone of height r and base radius r contains a ball of radius $r/(1 + \sqrt{2})$ and so $S_i K_{i-1}$ contains a ball of radius $(1/(2 + \sqrt{2}) > 1/4$. It follows that $S_i K_i$, which is a superset of $S_i K_{i-1}$, also contains such a ball.

If i is not a multiple of n, the algorithm estimates the variance along x_0 and we still have (20) for S_i. Let j be the largest multiple of n smaller than i. Then S_i can be viewed as S_j composed with a scaling along the x_0 axis. The projection of $S_j K_j$ orthogonal to x_0 contains a ball of radius $1/2$ and hence the projection of $S_i K_{i-1}$ contains such a ball as well. From (20), we get that $S_i K_{i-1}$ contains a cone of height $1/\sqrt{2}$ and therefore, as in the previous case, it contains a ball of radius $1/4$; so does $S_i K_i$.

Next, we have to bound $\mathsf{E}(|x - \bar{x}|^2)$. If i is a multiple of n, then using (19) and (20),

$$\begin{aligned} \mathsf{E}_{S_i K_{i-1}}(|x - \bar{x}|^2) &= \sum_{p=0}^{n} \mathsf{E}_{S_i K_{i-1}} \left((x - \bar{x})_p^2 \right) \\ &\leq \frac{3(n+1)}{2} \end{aligned}$$

where the inequality holds with high probability.

For any i (i.e., whether or not it is a multiple of n), by (20), with high probability,

$$\mathsf{E}_{S_iK_{i-1}}\left((x-\bar{x})_0^2\right) \leq \frac{3}{2}. \tag{21}$$

The support of S_iK_{i-1} along x_0 is $[0,u]$ where, by (20) and Lemma 3.4,

$$u \leq \sqrt{\frac{3}{2}n(n+2)} < \frac{3n}{2}.$$

Further, since K_{i-1} is a pencil, the cross-sectional area along x_0 is a nondecreasing function of x_0 (in its support) and so,

$$\mathsf{E}_{S_iK_{i-1}}((x-\bar{x})_0^2) \geq \frac{(u-\bar{x}_0)^2}{3}.$$

Therefore, using (21),

$$|u - \mathsf{E}_{S_iK_{i-1}}(x_0)| \leq \frac{9}{2}.$$

We also have $S_iK_i \subseteq 2^{1/n}S_iK_{i-1}$ and so the support of S_iK_i along x_0 is contained in $[0, 2^{1/n}u]$. Hence, there is some $\alpha \in [0,1]$ such that

$$\begin{aligned}
\mathsf{E}_{S_iK_i}\left((x-\bar{x})_0^2\right) &\leq \alpha \mathsf{E}_{S_iK_{i-1}}\left((x-\bar{x})_0^2\right) + (1-\alpha)\left(\frac{9}{2} + u(2^{\frac{1}{n}}-1)\right)^2 \\
&\leq \alpha\frac{3}{2} + (1-\alpha)\left(\frac{9}{2}+\frac{u}{n}\right)^2 \\
&\leq \alpha\frac{3}{2} + (1-\alpha)36 \leq 36.
\end{aligned}$$

It remains to bound the expected squared distance along coordinates other than x_0. Let j be the largest multiple of n smaller than or equal to i. Let the support of S_iK_{j-1} along x_0 be $[0,u]$ and that of S_iK_i be $[0,v]$. Note that orthogonal to x_0, S_i and S_j are the same transformation, and so,

$$\mathsf{E}_{S_iK_{j-1}}\left(\sum_{p=1}^{n}(x-\bar{x})_p^2\right) \leq \frac{3n}{2}.$$

Let u' be the rightmost nonempty cross-section of S_iK_{j-1} that is contained in a ball of radius $r_0 = 4\sqrt{n}\ln 2n$. Clearly, $u' \leq u$ and $v \leq 2^{1/n}u$. If $v \leq 3u'$, then every cross-section of S_iK_i is contained in a ball of radius $3r_0$ and so,

$$\mathsf{E}_{S_iK_i}\left(\sum_{p=1}^{n}(x-\bar{x})_p^2\right) \leq (3r_0)^2 = 36n\ln^2 2n.$$

Otherwise, $u' \leq 2u/3$ and so $|u-u'| \geq u/3$. Thus,

$$\lambda = \frac{|v-u|}{|v-u'|} \leq \frac{3}{4}.$$

Let the cross-sections of K_i at u', u and v be $B_{u'}$, B_u and B_v respectively. By the Brunn-Minkowski inequality,

$$\begin{aligned} \mathrm{vol}(B_u)^{\frac{1}{n}} &\geq \mathrm{vol}(\lambda B_{u'} + (1-\lambda)B_v)^{\frac{1}{n}} \\ &\geq \lambda \mathrm{vol}(B_{u'})^{\frac{1}{n}} + (1-\lambda)\mathrm{vol}(B_v)^{\frac{1}{n}}. \end{aligned} \tag{22}$$

Further, by the choice of r_0, applying Lemma 3.3 to $S_i K_{j-1}$,

$$\mathrm{vol}(B_u) \leq \mathrm{vol}(B_{u'})(1 + \frac{1}{8n^2}).$$

(Otherwise the volume of $S_i K_{j-1}$ between u' and u is too large.) Therefore, by (22),

$$\mathrm{vol}(B_v) \leq \mathrm{vol}(B_{u'})(1 + \frac{2}{n^2}).$$

So, for any cross-section of $S_i K_i$ between u and v, the fraction of the volume that lies outside a ball of radius r_0 is at most $\frac{2}{n^2}$. Now, by Lemma 3.4, each cross-section of $S_i K_{j-1}$ is contained in a ball of radius $2n$ and so B_u is contained in such a ball. Thus, B_v is contained in a ball of radius $4n$. Therefore,

$$\mathsf{E}_{S_i K_i}\left(\sum_{p=1}^{n}(x-\bar{x})_0^2\right) \leq r_0^2 + \frac{2}{n^2}(4n)^2 \leq 16n(\ln^2 2n + 2).$$

This completes the proof of the lemma. □

6 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. All previous algorithms used $\Omega(n^2)$ samples.

2. If we use $a_{i+1} = a_i(1 - 1/t)$ then the number of phases is $m = O(t \ln(n/\varepsilon))$. By the remark after Lemma 4.1 and a similar analysis of cumulative sampling error, for any $t = \Omega(\sqrt{n})$, we need only $O(mn/t^2\varepsilon^2)$ samples in each phase. So e.g., with $t = n$, we have $O(n \ln(n/\varepsilon))$ phases and $O(\ln(n/\varepsilon)/\varepsilon^2)$ samples in each phase.

3. It is a natural next step to extend this method to integration of logconcave functions. The fundamental Lemma 3.2 can be extended to this case, but certain technical results we used from [18] 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.

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 discussions and for his help in simplifying the proof of Lemma 3.2.

References

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

[2] I. Bárány and Z. Furedi: Computing the Volume is Difficult. Discrete & Computational Geometry 2, 1987, 319-326.

[3] D. Bertsimas and S. Vempala: Solving convex programs by random walks. *Proc. STOC*, 2002, 109-115. To appear in *J. ACM*.

[4] A. Dinghas: Über eine Klasse superadditiver Mengenfunktionale von Brunn–Minkowski–Lusternik-schem Typus, *Math. Zeitschr.* **68** (1957), 111–125.

[5] M. Dyer and A. Frieze: Computing the volume of convex bodies: a case where randomness provably helps. *Proc. AMS Symposium on Probabilistic Combinatorics and Its Applications*, (1991), 123-170.

[6] 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.

[7] G. Elekes: A Geometric Inequality and the Complexity of Computing Volume. Discrete & Computational Geometry 1, 1986, 289-292.

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

[9] 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.

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

[11] L. Leindler: On a certain converse of Hölder's Inequality II, *Acta Sci. Math. Szeged* 33 (1972), 217–223.

[12] L. Lovász: How to compute the volume? *Jber. d. Dt. Math.-Verein, Jubiläumstagung 1990*, B. G. Teubner, Stuttgart, 138–151.

[13] L. Lovász and M. Simonovits: Mixing rate of Markov chains, an isoperimetric inequality, and computing the volume, *Proc. 31st IEEE Annual Symp. on Found. of Comp. Sci.* (1990), 482–491.

[14] 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.

[15] L. Lovász and M. Simonovits: Random walks in a convex body and an improved volume algorithm, *Random Structures and Alg.* **4** (1993), 359–412.

[16] 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

[17] 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

[18] L. Lovász and S. Vempala: Hit-and-run from a corner. To appear in *Proc. of STOC*, 2004.

[19] A. Prékopa: Logarithmic concave measures and functions, *Acta Sci. Math. Szeged* 34 (1973), 335–343.

[20] A. Prékopa: On logarithmic concave measures with applications to stochastic programming, *Acta Sci. Math. Szeged* 32 (1973), 301–316.

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