Fast Algorithms for Logconcave Functions: Sampling, Rounding, Integration and Optimization

László Lovász
Microsoft Research
lovasz@microsoft.com

Santosh Vempala *
Georgia Tech and MIT
vempala@cc.gatech.edu

Abstract

We prove that the hit-and-run random walk is rapidly mixing for an arbitrary logconcave distribution starting from any point in the support. This extends the work of [26], where this was shown for an important special case, and settles the main conjecture formulated there. From this result, we derive asymptotically faster algorithms in the general oracle model for sampling, rounding, integration and maximization of logconcave functions, improving or generalizing the main results of [24, 25, 1] and [16] respectively. The algorithms for integration and optimization both use sampling and are surprisingly similar.

1 Introduction

Given a real-valued function f in $\mathbb{R}^n$, computing the integral of f and finding a point that maximizes f are two fundamental algorithmic problems. They include as special cases, the famous problems of computing the volume of a convex body (here f is the indicator function of the convex body) and minimizing a linear function over a convex set (here f is equal to the linear function over the convex set and infinity outside). In this general setting, the complexity of an algorithm can be measured by the number of function evaluations ("membership queries"). Both problems are computationally intractable in general [7, 2, 5, 17] and the question of understanding the set of functions for which they are solvable in polynomial time has received much attention over the past few decades.

In a breakthrough paper, Dyer, Frieze and Kannan [6] gave a polynomial-time algorithm ($O^*(n^{23})$) for the special case of computing the volume of a convex body. Many improvements followed [1, 19, 21, 23, 5, 14, 25], the most recent reducing the complexity to $O^*(n^4)$. Applegate and Kannan [1] gave an algorithm for the integration of logconcave functions (with mild smoothness assumptions), a common generalization of convex bodies and Gaussians. The driving idea behind all these results is random sampling by geometric random walks. Unlike volume computation, there has not been much improvement in the complexity of integration and the current best for general logconcave functions is still $O^*(n^{10})$ from [1] (the O^* notation suppresses the dependence on error parameters and logarithmic terms).

For minimizing a convex function over a convex set (equivalent to maximizing a logconcave function), the original solution was the deterministic ellipsoid method [10]. The current best algorithm is based on sampling by a random walk and its complexity is $O^*(n^5)$ [3] (with a separation oracle, the complexity is $O^*(n)$ [30, 3]). This has recently been improved to $O^*(n^{4.5})$ for minimizing a linear function over a convex set [16].

The class of functions for which either integration or optimization can be achieved in (randomized) polynomial time is currently determined by *efficient sampling* from the distribution whose density is proportional to the given function. In [24], it is shown that arbitrary logconcave functions can be sampled using the ball walk or the hit-and-run walk. However, the walks are proved to be rapidly mixing only from a *warm* start, i.e., a distribution that is already close to the target. This requirement leads to a considerable overhead in the algorithm (and, of course, in the analysis). For the ball walk, such a requirement is unavoidable in general—the mixing rate depends polynomially on the "distance" between the start and target distributions (see Section 1.2).

On the other hand, the hit-and-run algorithm, which seems to be the fastest sampling algorithm in practice, is conjectured to be rapid-mixing from any starting distribution (including one that is concentrated on a single point); more precisely, the dependence on the distance to the stationary distribution could be *logarithmic*. An important step towards proving this conjecture was taken in [26] where it was proved for the special case when the target distribution is an exponential function over a convex body (which includes the uniform density). This result plays a key role in the latest volume algorithm [25] and also in a faster algorithm for minimizing a linear function over a convex set [16]. The major hurdle in extending these algorithms to integration and optimization of logconcave functions is the lack of a provably rapid-mixing random walk with a similar mild (logarithmic) dependence on the start.

*Supported in part by a Guggenheim fellowship.

1.1 Computational model

First, let us have a brief discussion of the case of convex bodies. One usual definition of mixing time is to take the worst starting point, but we cannot use this here. Suppose that we want to sample from the unit n-cube and we start at a vertex or infinitesimally close to it. Then, for each of the algorithms considered so far, it takes about 2^n time before the orthant in which a nontrivial step can be made is recognized, since our only access to the body is a membership oracle.

One consequence of this is that if we want polynomial bounds in the dimension n, we have to either provide information about the starting point, say by taking into account the distance of the point to the boundary. Else, we can start from a random point from a distribution which is sufficiently spread out. A strong version of this second approach is *warm start*, when we assume the density of the starting distribution relative to the stationary distribution is at most 2.

A theorem of Aldous and Diaconis guarantees that if the distance of the distribution after m_0 steps from the stationary distribution *from any starting point* is less than $1/4$, then after m steps it is less than $2^{-m/m_0}$; but this gives very poor bounds here because of "bad" starting points. So polynomiality in $|\log\varepsilon|$ (where ε is the distance we want) is an issue here.

To be more precise, a *convex body sampling problem* is given by

(CS1) a convex body $K \subseteq \mathbb{R}^n$, which is given by a membership oracle, together with two "guarantees" r and R that it contains a ball with radius r and it is contained in a ball with radius R;

(CS2) a starting point $a_0 \in K$, together with a "guarantee" $d > 0$ such that $B(a_0, d) \subseteq K$;

(CS3) an error bound $\varepsilon > 0$, describing how close the sample point's distribution must be to the uniform distribution on K (in total variation distance).

Instead of (CS2), we could have

(CS2') a random starting point $a_0 \in K$, together with a "guarantee" $M \geq 1$ such that the density of the distribution σ of a_0 is bounded by M (relative to the uniform density on K)

or

(CS2") a random starting point $a_0 \in K$, together with a parameter $M \geq 1$ such that the density of the distribution σ of a_0 is bounded by M except in a subset S with $\sigma(S) \leq \varepsilon/2$.

It turns out the (CS2") is the most useful version for applications.

More generally, we consider the problem of sampling from a logconcave distribution. To state the problem formally, we consider a logconcave function $f : \mathbb{R}^n \to \mathbb{R}+$; it will be convenient to assume that f has a bounded support, so that there s a convex body K such that $f = 0$ outside K. We consider the distribution π_f defined by $\pi_f(X) = \left(\int_X f\right)/\left(\int_{\mathbb{R}^n} f\right)$. So f is proportional to the density function of the distribution, but not necessarily equal, i.e., we don't assume that $\int f = 1$. We'll need the level sets $L(a) = \{x \in \mathbb{R}^n : f(x) \geq a\}$. The *centroid* of π_f is $z_f = \int x \, d\pi_f(x) = \mathsf{E}(x)$, and the *variance* of π_f is $\mathsf{Var}(\pi_f) = \int \|x - z_f\|^2 \, d\pi_f = \mathsf{E}(\|x - z_f\|^2)$.

This sampling problem is given by:

(LS1) an oracle that evaluates f at any given point, together with "guarantees" that the distribution is neither too spread out nor to concentrated; this will be given by $r, R > 0$ such that if $\pi_f(L(c)) \geq 1/8$ then $L(c)$ contains a ball of radius r, and the variance of π_f is at most R^2.

(LS2) a starting point $a_0 \in \mathbb{R}^n$, together with a "guarantee" $d > 0$ such that a_0 is at distance d from the boundary of the level set $L(f(a_0)/2)$ and another guarantee $\beta > 0$ such that $f(a_0) \geq \beta^n \max f$.

(LS3) an error bound $\varepsilon > 0$, describing how close the sample point's distribution must be to the target distribution π_f (in total variation distance).

Again, we could replace (LS2) by the assumption that we have

(LS2') a random starting point a_0 from a distribution σ, and a "guarantee" $M \geq 1$ such that $d\sigma/d\pi_f \leq M$

or even more generally,

(LS2") a random starting point a_0 from a distribution σ, and a "guarantee" $M \geq 1$ such that $d\sigma/d\pi_f \leq M$ except for a set S with $\sigma(S) \leq \varepsilon$.

The sampling algorithm we analyze is the hit-and-run walk in $\mathbb{R}^n$, which we start at a_1 and stop after m steps (see section 3 for the definition). Let σ^m denote the distribution of the last point. We say that the walk is *rapidly mixing* if for some m polynomial in n, $\log R$, $\log r$, $\log\varepsilon$, $\log d$ and $\log\beta$ (or $\log M$), $d_{\rm tv}(\pi_f, \sigma^m) \leq \varepsilon$.

1.2 Results

In this paper, we first show that hit-and-run applied to any logconcave function mixes rapidly. The mixing time is $O^*(n^3)$ after appropriate normalization (to be discussed shortly). It is worth noting that this is the first random walk proven to be rapidly mixing (in the sense above) even for the special case of convex bodies. (Strictly speaking, the other well-known walks, namely the ball walk and the lattice walk ([8, 9]), are not rapid-mixing.)

Theorem 1.1 *Let f be a logconcave function in $\mathbb{R}^n$, given in the sense of (LS1), (LS2") and (LS3). Then for*

$$m > 10^{30} \frac{n^2 R^2}{r^2} \ln^2 \frac{MnR}{r\varepsilon} \ln^3 \frac{M}{\varepsilon},$$

the total variation distance of σ^m and π_f is less than 2ε.

Remarks: 1. If instead of a bound on the moment $\mathsf{E}(|x - z_f|^2)$ we have the stronger condition that the support of f is contained in a ball of radius R, then the bound on the mixing time is smaller by a factor of $\ln^2(M/\varepsilon)$.

2. The condition (LS2") captures the setting when L_2 distance (see Section 2) between the start and target densities is bounded.

3. This result improves on the main theorem of [24] by reducing the dependence on M and ε from $(M/\varepsilon)^4$ to polylogarithmic. For the ball walk, some polynomial

dependence on M is unavoidable. (consider e.g., a starting distribution that is uniform in the set of points within $\delta/2$ of the apex of a rotational cone, where δ is the radius of the ball walk).

To analyze hit-and-run starting at a single point, we take one step and then apply the theorem above to the distribution obtained.

Corollary 1.2 *Let f be a logconcave function in $\mathbb{R}^n$, given in the sense of (LS1), (LS2) and (LS3). Then for*

$$m > 10^{31} \frac{n^3 R^2}{r^2} \ln^5 \frac{nR^2}{\varepsilon r d\beta},$$

the total variation distance of σ^m and π_f is less than ε.

The bound above does not imply rapid-mixing, because the number of steps is polynomial in R/r, rather than in its logarithm. We call a logconcave function *well-rounded*, if $R/r = O(\sqrt{n})$. In the rest of this introduction, we assume that our logconcave functions are well-rounded. In this case, our algorithms are polynomial, and the (perhaps most interesting) dependence on the dimension is $O^*(n^3)$. Every logconcave function can be brought to a well-rounded position by an affine transformation of the space in polynomial time (we'll return to this shortly).

With this efficient sampling algorithm at hand, we turn to the problems of integration of logconcave functions, perhaps the most important problem considered here, at least from a practical viewpoint. The idea for integration, first suggested in [22], can be viewed as a generalized version of the method called *simulated annealing* in a continuous setting. We consider a sequence of functions of the form $f(x)^{1/T}$ where T, the "temperature", starts very high (which means the corresponding distribution is nearly uniform over the support of f) and is reduced to 1. We will see shortly that the *same* idea is also suitable for maximization of logconcave functions.

Theorem 1.3 *Let f be a well-rounded logconcave function given by a sampling oracle. Given $\varepsilon, \delta > 0$, we can compute a number A such that with probability at least $1 - \delta$,*

$$(1 - \varepsilon) \int f \le A \le (1 + \varepsilon) \int f$$

and the number of oracle calls is

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

In [25], it was shown that a convex body can be put in near-isotropic position (which implies well-roundedness, see Section 2) in $O^*(n^4)$ steps. In Section 6, we give an algorithm to put an arbitrary logconcave function in near-isotropic position in $O^*(n^4)$ steps. However, this algorithm (as previous algorithms) requires the knowledge of a point where f is (approximately) maximized (a stronger version of (LS2)). In many cases, such a point is much easier to find. If one has a *separation oracle* for f, which, given a point x, returns a hyperplane that separates x from points that attain the maximum (e.g. a tangent plane to the level set $L(f(x))$), then the maximization problem can be solved using $O^*(n)$ oracle calls [3, 30]. This gives an integration algorithm of complexity $O^*(n^4)$ for arbitrary logconcave functions, improving on the $O^*(n^{10})$ algorithm of Applegate and Kannan.

It is, however, possible that the only access to f is via an oracle that evaluates f at a given point. In this setting, the maximization problem seems more difficult. Our last result is an algorithm for this problem. The idea for maximization, described in [16] and earlier for a more restricted setting in [15], is again simulated annealing: we consider a sequence of functions $f(x)^{1/T}$ where T starts very high (so the distribution is close to uniform) and is reduced till the distribution is concentrated sufficiently close to points that maximize f (so, unlike integration, we do not stop when $T = 1$ but proceed further till T is sufficiently close to zero). The resulting algorithm can be viewed as an interior-point method that requires only $O^*(\sqrt{n})$ phases. By using $f(x) = e^{-g(x)}$, we get an algorithm for minimizing any convex function g over a convex body.

Theorem 1.4 *For any well-rounded logconcave function f, given $\varepsilon, \delta > 0$ and a point x_0 with $f(x_0) \ge \beta^n \max f$, we can find a point x in $O^*(n^{4.5})$ oracle calls such that with probability at least $1 - \delta$,*

$$f(x) \ge (1 - \varepsilon) \max f$$

and the dependence on ε, δ and β is bounded by a polynomial in $\ln(1/\varepsilon\delta\beta)$.

This improves on the previous best algorithm [3] by a factor of $\sqrt{n}$.

2 Preliminaries

We measure the distance between two distributions σ and π in two ways, by the *total variation distance*

$$d_{tv}(\sigma, \pi) = \frac{1}{2} \int_{\mathbb{R}^n} |\sigma(x) - \pi(x)|\, dx,$$

and the L_2 *distance*

$$d_2(\sigma, \pi) = \int_{\mathbb{R}^n} \frac{\sigma(x)}{\pi(x)}\, d\sigma(x).$$

A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is said to be *logconcave* if $f(\alpha x + (1 - \alpha)y) \ge f(x)^\alpha f(y)^{1-\alpha}$ for every $x, y \in \mathbb{R}^n$ and $0 \le \alpha \le 1$. Gaussians, exponentials and indicator functions over convex bodies are all logconcave. The product, minimum and convolution of two logconcave functions is also logconcave [4, 27, 18]. The last property implies that any marginal of a logconcave function is also logconcave. We will use several known properties of logconcave functions. The following property proved in [16] (a variant of a lemma from [25]) will be used multiple times.

Lemma 2.1 ([16]) *Let f be a logconcave function in $\mathbb{R}^n$. For $a > 0$, let*

$$Z(a) = \int_{\mathbb{R}^n} f(x)^a \, dx.$$

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

The level sets of a function f are denoted by $L_f(t) = \{x \,|\, f(x) \geq t\}$ or just $L(t)$ when the context is clear. For a logconcave f, the level sets are all convex.

A density function f is said to be in *isotropic* position if $\mathsf{E}_f(x) = 0$ and $\mathsf{E}_f(xx^T) = I$. The second condition is equivalent to saying that

$$\int_{\mathbb{R}^n} (v^T x)^2 f(x) \, dx = 1$$

for any unit vector $v \in \mathbb{R}^n$. Similarly, f is C-isotropic if

$$\frac{1}{C} \leq \int_{\mathbb{R}^n} (v^T(x - z_f))^2 f(x) \, d(x) \leq C$$

for any unit vector $v \in \mathbb{R}^n$. For any function, there is an affine transformation that puts it in isotropic position. An approximation to this can be computed from a random sample as implied by combining a lemma from [28] with a moment bound from [24] (see e.g., Corollary A.2 in [16]). We quote it below as a theorem.

Lemma 2.2 *Let f be a logconcave function in $\mathbb{R}^n$ that is not concentrated on a subspace and let $X^1, \ldots X^k$ be independent random points from the distribution π_f. There is a constant C_0 such that if $k > C_0 t^3 \ln n$ then the transformation $g(x) = T^{-1/2}x$ where*

$$\bar{X} = \frac{1}{k}\sum_{i=1}^{k} X^i, \quad T = \frac{1}{k}\sum_{i=1}^{k}(X^i - \bar{X})(X^i - \bar{X})^T$$

puts f in 2-isotropic position with probability at least $1 - 1/2^t$.

A weaker normalization will also be useful. We say that a function f is *C-round* if $R/r \leq C\sqrt{n}$ where $R^2 = \mathsf{E}_f(|X - z_f|^2)$ and the level set of f of measure $1/8$ contains a ball of radius r. When C is a constant, we say that f is *well-rounded*. For an isotropic function f, we have $\mathsf{E}_f(|X|^2) = n$. In [24], it is shown that if f is also logconcave, then any level set L contains a ball of radius $\pi_f(L)/e$. Thus, if a logconcave function is isotropic than it is also $(8e)$-round.

3 Analysis of the hit-and-run walk

The hit-and-run walk for an integrable, nonnegative function f in $\mathbb{R}^n$ is defined as follows:

- Pick a uniformly distributed random line ℓ through the current point.
- Move to a random point y along the line ℓ chosen with density proportional to f (restricted to the line).

It is well-known that the stationary distribution of this walk is π_f. Our goal is to bound the rate of convergence. We will use the conductance technique of Jerrum and Sinclair [11] as extended to the continuous setting by Lovász and Simonovits [23]. The state space of hit-and-run is the support K of f. For any measurable subset $S \subseteq K$ and $x \in K$, we denote by $P_x(S)$ the probability that a step from x goes to S. For $0 < \pi_f(S) < 1$, the conductance $\phi(S)$ is defined as

$$\phi(S) = \frac{\int_{x\in S} P_x(K \setminus S) \, d\pi_f}{\min\{\pi_f(S), \pi_f(K \setminus S)\}}.$$

In words, this is the probability of going from S to its complement in one step, given that we start randomly on the side of smaller measure. Then, if ϕ is the minimum conductance over all measurable subsets, $O(1/\phi^2)$ is a bound on the mixing time (roughly speaking, the number of steps required to halve the distance to the stationary distribution).

At a high level, the analysis has two parts: (i) "large" subsets (under the measure defined by f) have large boundaries and (ii) points that are "near" the boundary are likely to cross from a subset to its complement in one step. The first part will be given by an *isoperimetric inequality*. In fact, we will use the weighted inequality developed in [26], quoted below. This geometric inequality is independent of any particular random walk. It is the second part that is quite different for different random walks. It consists of connecting appropriate notions of geometric distance and probabilistic distance by a careful analysis of single steps of the random walk.

The isoperimetric inequality below uses the cross-ratio distance defined as follows. For two points u, v in a convex body K, let p, q be the endpoints of the line through u, v so that they appear in the order p, u, v, q. Then the cross-ratio distance is

$$d_K(u, v) = \frac{|u - v||p - q|}{|p - u||v - q|}.$$

Theorem 3.1 ([26]) *Let K be a convex body in $\mathbb{R}^n$. Let $f : K \to \mathbb{R}_+$ be a logconcave function, π_f be the corresponding distribution and $h : K \to \mathbb{R}_+$, an arbitrary function. Let S_1, S_2, S_3 be any partition of K into measurable sets. Suppose that for any pair of points $u \in S_1$ and $v \in S_2$ and any point x on the chord of K through u and v,*

$$h(x) \leq \frac{1}{3}\min(1, d_K(u, v)).$$

Then

$$\pi_f(S_3) \geq \mathsf{E}_f(h(x)) \min\{\pi_f(S_1), \pi_f(S_2)\}.$$

For a function f in $\mathbb{R}^n$, let $\mu_{\ell,f}$ be the measure induced by f on the line ℓ. For two points $u, v \in \mathbb{R}^n$, let ℓ be the line through them, ℓ^- be the semi-line starting at u

not containing v and ℓ^+ be the semiline starting at v not containing u. Then the f-distance is

$$d_f(u,v) = \frac{\mu_{\ell,f}([u,v])\mu_{\ell,f}(\ell)}{\mu_{\ell,f}(\ell^-)\mu_{\ell,f}(\ell^+)}$$

3.1 The smoothness of single steps

As in [20, 26], at a point x, we define the *step-size* $F(x)$ by

$$\mathsf{P}\left(|x-y| \le F(x)\right) = \frac{1}{8},$$

where y is a random step from x. Next, we define

$$\lambda(x,t) = \frac{\text{vol}((x+tB)\cap L((3/4)f(x)))}{\text{vol}(tB)}$$

and

$$s(x) = \sup\{t \in \mathbb{R}_+ : \lambda(x,t) \ge 63/64\}.$$

Finally, we define $\alpha(x)$ to be the smallest $t \ge 3$ for which a hit-and-run step y from x satisfies $\mathsf{P}(f(y) \ge tf(x)) \le 1/16$. The following lemma was proved in [24].

Lemma 3.2 ([24])

$$\pi_f(u : \alpha(u) \ge t) \le \frac{16}{t}.$$

The next two lemmas are from [26].

Lemma 3.3 ([26])

$$F(x) \ge \frac{s(x)}{64}.$$

Lemma 3.4 ([26]) *Let f be any logconcave function such that the level set of f of measure $1/8$ contains a ball of radius r. Then*

$$\mathsf{E}_f(s(x)) \ge \frac{r}{2^{10}\sqrt{n}}.$$

Next we prove a new lemma that will be crucial in extending the analysis of hit-and-run.

Lemma 3.5 *Let $u, v \in K$, the support of a logconcave function f and let $x \in [u,v]$. If $s(x) \ge 4c|u-v|\sqrt{n}$ for some $c \ge 1$, then*

$$d_f(u,v) \le \frac{1}{c} \quad \textit{and} \quad |u-v| \le \frac{\max\{s(u), s(v)\}}{c\sqrt{n}}.$$

Proof. First suppose $d_f(u,v) \ge 1/c$. Let y, y' be the points on the line $\ell(u,v)$ at distance $4c|u-v|$ from x. From the definition of d_f and the logconcavity of f, $\min\{f(y), f(y')\} \le f(x)/2$. Assume $f(y) \le f(x)/2$. Let H be a supporting plane at y of the level set $L(f(y))$. In the ball of radius $4c|u-v|\sqrt{n}$ around x, the halfspace bounded by H not containing x cuts off at least $1/16$th of the volume of the ball; further, each point in the separated subset has function value at most $f(y) \le f(x)/2$. Therefore, $s(x) < 4c|u-v|\sqrt{n}$.

Suppose that $|u-v| \ge \max\{s(u), s(v)\}/c\sqrt{n}$. We may suppose that $f(u) \le f(x)$ (one of u, v has function value at most $f(x)$). The level set $L(3f(u)/4)$ misses $1/64$ of the ball of radius $s(u)$ around u. Using Lemma 4.4 in [24], it misses at least $3/4$ of the ball of radius $4c|u-v|\sqrt{n} > 4s(u)$ around u. Now consider balls of radius $4c|u-v|\sqrt{n}$ around u and x. These balls overlap in at least $1/2$ of their volume. Therefore, the level set $L(3f(u)/4)$ misses at least $1/4$ of the ball around x; since $f(u) \le f(x)$, the level set $L(3f(x)/4) \subseteq L(3f(u)/4)$ and so $L(3f(x)/4)$ also misses at least $1/4$ of the ball around x. This contradicts the assumption that $s(x) \ge 4c|u-v|\sqrt{n}$. □

3.2 Conductance and mixing time

For a point $u \in K$, let P_u be the distribution obtained by taking one hit-and-run step from u. Let $\mu_f(u,x)$ be the integral of f along the line through u and x. Then,

$$P_u(A) = \frac{2}{n\pi_n}\int_A \frac{f(x)\,dx}{\mu_f(u,x)|x-u|^{n-1}}. \qquad (1)$$

The next lemma from [26] connects geometric distance with probabilistic distance.

Lemma 3.6 ([26]) *Let $u, v \in K$. Suppose that*

$$d_f(u,v) < \frac{1}{128\ln(3+\alpha(u))}$$

and

$$|u-v| < \frac{1}{4\sqrt{n}}\max\{F(u), F(v)\}.$$

Then $d_{tv}(P_u, P_v) < 1 - 1/500$.

We are now ready to prove the main theorem bounding the conductance.

Theorem 3.7 *Let f be a logconcave function in $\mathbb{R}^n$ such that the level set of measure $1/8$ contains a unit ball and the support K has diameter D. Then for any subset S, with $\pi_f(S) = p \le 1/2$, the conductance of hit-and-run satisfies*

$$\phi(S) \ge \frac{1}{10^{13}nD\ln(\frac{nD}{p})}.$$

Proof. Let $K = S_1 \cup S_2$ be a partition into measurable sets, where $S_1 = S$ and $p = \pi_f(S_1) \le \pi_f(S_2)$. We will prove that

$$\int_{S_1} P_x(S_2)\,dx \quad \ge \quad \frac{1}{10^{13}nD\ln\frac{nD}{p}}\pi_f(S_1) \qquad (2)$$

Consider the points that are deep inside these sets:

$$S_1' = \left\{x \in S_1 : P_x(S_2) < \frac{1}{1000}\right\}$$

and

$$S_2' = \left\{x \in S_2 : P_x(S_1) < \frac{1}{1000}\right\}.$$

Let S_3' be the rest i.e., $S_3' = K \setminus S_1' \setminus S_2'$.

Suppose $\pi_f(S_1') < \pi_f(S_1)/2$. Then

$$\int_{S_1} P_x(S_2)\,dx \geq \frac{1}{1000}\pi_f(S_1 \setminus S_1') \geq \frac{1}{2000}\pi_f(S_1)$$

which proves (2). So we can assume that $\pi_f(S_1') \geq \pi_f(S_1)/2$ and similarly $\pi_f(S_2') \geq \pi_f(S_2)/2$.

Next, define the exceptional subset W as set of points u for which $\alpha(u)$ is very large.

$$W = \left\{u \in S : \alpha(u) \geq \frac{2^{27}nD}{p}\right\}.$$

By Lemma 3.2, $\pi_f(W) \leq p/2^{23}nD$. Now, for any $u \in S_1' \setminus W$ and $v \in S_2' \setminus W$,

$$d_{tv}(P_u, P_v) \geq 1 - P_u(S_2) - P_v(S_1) > 1 - \frac{1}{500}.$$

Thus, by Lemma 3.6,

$$d_f(u,v) \geq \frac{1}{128\ln(3+\alpha(u))} \geq \frac{1}{2^{12}\ln\frac{nD}{p}}$$

or

$$|u-v| \geq \frac{1}{4\sqrt{n}}\max\{F(u), F(v)\}.$$

By Lemma 3.3, the latter implies that

$$|u-v| \geq \frac{1}{2^8\sqrt{n}}\max\{s(u), s(v)\}.$$

Now applying Lemma 3.5, in either case, for any point $x \in [u,v]$, we have

$$\begin{aligned} s(x) &\leq 2^{14}\ln\left(\frac{nD}{p}\right)|u-v|\sqrt{n} \\ &\leq 2^{14}\ln\left(\frac{nD}{p}\right)d_K(u,v)D\sqrt{n}. \end{aligned}$$

To apply Theorem 3.1, we define

$$h(x) = \frac{s(x)}{2^{16}D\sqrt{n}\ln\frac{nD}{p}}$$

and consider the partition $S_1' \setminus W$, $S_2' \setminus W$ and the rest. Clearly, for any $u \in S_1' \setminus W, v \in S_2' \setminus W$ and $x \in [u,v]$, we have $h(x) \leq d_K(u,v)/3$. Thus,

$$\begin{aligned} \pi_f(S_3') &\geq \mathsf{E}_f(h)\pi_f(S_1' \setminus W)\pi_f(S_2' \setminus W) - \pi_f(W) \\ &\geq \frac{1}{2^{29}nD\ln\frac{nD}{p}}\pi_f(S_1). \end{aligned}$$

Here we have used Lemma 3.4 and the bound on $\pi_f(W)$. Therefore,

$$\begin{aligned} \int_{S_1} P_x(S_2)\,dx &\geq \frac{1}{2}\cdot\frac{1}{1000}\pi_f(S_3') \\ &\geq \frac{1}{10^{13}nD\ln\frac{nD}{p}}\pi_f(S_1) \end{aligned}$$

which again proves (2) □

3.3 Mixing time

When the support of f is bounded by a ball of radius R, the bound on the mixing time in Theorem 1.1 follows by applying Corollary 1.6 in [23] with $p = \varepsilon/2M$ and $D = 2R/r$ in the expression for conductance.

Now let $\mathsf{E}_f(\|x - z_f\|^2) \leq R^2$. We consider the restriction of f to the ball of radius $R\ln(4e/p)$ around the centroid z_f. By Lemma 5.17 in [24], the measure of f outside this ball is at most $p/4$. In the proof of the conductance bound, we can consider the restriction of f to this set. In the bound on the conductance for a set of measure p, the diameter D is effectively replaced by $R\ln(4e/p)$, i.e., for any subset of measure p, the conductance is at least

$$\phi(p) \geq \frac{cr}{nR\ln(nR/rp)\ln(4e/p)}$$

where c is a fixed constant. The bound on mixing follows by applying Corollary 1.6 in [23] with $p = \varepsilon/2M$.

4 Integration

The algorithm can be viewed as a generalized version of simulated annealing. To integrate f, we start with the constant function over K, the support of f, and slowly transform it into f via a sequence of $m = O^*(\sqrt{n})$ log-concave functions. In each step, we compute the ratio of consecutive integrals by sampling. Multiplying all of these and the volume of K, we get our answer.

In the description below, we assume that f is well-rounded. We restrict f to a ball of radius $O(\sqrt{n}\ln(1/\varepsilon))$ and to the larger of the level sets $L(f(x_0)/2)$ and $L(\max f/e^{-2n-2\ln(1/\varepsilon)})$. By Lemma 6.13 of [24], this restriction removes a negligible part of the integral of f. The algorithm uses an upper bound B on the quantity $\ln(\max f/\min f)$. We can set $B = 2n + 2\ln(1/\varepsilon) + n\ln(1/\beta)$ where the initial point x_0 satisfies $f(x_0) \geq \beta^n \max f$.

Integration algorithm:

I0. Set $m = \lceil \sqrt{n} \ln B \rceil, k = \frac{512}{\varepsilon^2} \sqrt{n} \ln B$. Let K be the support of f and f_0 be the indicator function of K. For $i = 1, \ldots, m-1$, let

$$a_i = \frac{1}{B} \left(1 + \frac{1}{\sqrt{n}}\right)^i \text{ and } f_i(x) = f(x)^{a_i}$$

and $f_m(x) = f(x)$.

I1. Let $W_0 = \text{vol}(K)$ or an estimate of $\text{vol}(K)$ using the algorithm of [25]. Also, let $X_0^1, \ldots, X_0^k$ be independent uniform random points from K.

I2. For $i = 1, \ldots, m$, do the following.

— Run the sampler k times with target density proportional to f_{i-1} and starting points $X_{i-1}^1, \ldots X_{i-1}^k$ to get independent random points $X_i^1, \ldots, X_i^k$.

— Using these points, compute

$$W_i = \frac{1}{k} \sum_{j=1}^{k} f(X_i^j)^{a_i - a_{i-1}}.$$

I3. Return $W = W_0 W_1 \ldots W_m$.

To prove Theorem 1.3 we will show that (i) the estimate is accurate (Lemma 4.2) and (ii) use Theorem 1.1 to ensure that each sample is obtained using $O(n^3 \ln^5(n/\varepsilon))$ steps; this is implied by (a) samples from each distribution provide a good start to sampling the next distribution (Lemma 4.3) and (b) the density being sampled remains $O(\ln(1/\varepsilon))$-round (Lemma 4.4). The next lemma is analogous to Lemma 4.1 in [25] with similar proof. In this section, we let $Z(a)$ denote $\int_{\mathbb{R}^n} f(x)^a \, dx$.

Lemma 4.1 *Let X be a random point with density proportional to f_{i-1} and define $Y = f_i(X)/f_{i-1}(X)$. Then $\mathsf{E}(Y) = Z(a_i)/Z(a_{i-1})$ and $\mathsf{E}(Y^2) \le e\mathsf{E}(Y)^2$.*

Proof. The first part is straightforward. For the second we have

$$\begin{aligned}\mathsf{E}(Y^2) &= \int f(X)^{2a_i - 2a_{i-1}} \left(\frac{f(X)^{a_{i-1}}}{\int f(x)^{a_{i-1}} \, dx} \right) dX \\ &= \frac{Z(2a_i - a_{i-1})}{Z(a_{i-1})}.\end{aligned}$$

Now by Lemma 2.1, $a^n Z(a)$ is logconcave and so

$$(2a_i - a_{i-1})^n Z(2a_i - a_{i-1}) a_i^n Z(a_{i-1}) \le a_i^{2n} Z(a).$$

Using this, and the fact that $a_i = a_{i-1}(1 + 1/\sqrt{n})$, we have

$$\begin{aligned}\frac{\mathsf{E}(Y^2)}{\mathsf{E}(Y)^2} &= \frac{Z(2a_i - a_{i-1})Z(a_{i-1})}{Z(a_i)^2} \\ &\le \left(\frac{a_i^2}{(2a_i - a_{i-1})a_{i-1}} \right)^n \\ &= \left(\frac{\left(1 + \frac{1}{\sqrt{n}}\right)^2}{\left(1 + \frac{2}{\sqrt{n}}\right)} \right)^n \le \left(1 + \frac{1}{n}\right)^n \le e.\end{aligned}$$

□

Lemma 4.2 *With probability at least $3/4$, the estimate W satisfies*

$$(1 - \varepsilon) \int f \le W \le (1 + \varepsilon) \int f.$$

Proof. For $i = 1, \ldots, m$, let $R_i = Z(a_i)/Z(a_{i-1})$ and for $j = 1, \ldots, k$, let $Y_i^j = f(X_i^j)^{a_i - a_{i-1}}$. Then $\mathsf{E}(Y_i^j) = R_i$ and by Lemma 4.1, $\mathsf{E}((Y_i^j)^2) \le eR_i^2$. Also, $\mathsf{E}(W_i) = R_i$ and *assuming* the Y_j^i's are all independent,

$$\mathsf{E}(W_i^2) \le \left(1 + \frac{e-1}{k}\right) R_i^2.$$

Finally, $\mathsf{E}(W) = \text{vol}(K)\Pi_{i=1}^m \mathsf{E}(W_i) = \int_{\mathbb{R}^n} f(x) \, dx$ and

$$\mathsf{E}(W^2) \le \left(1 + \frac{e-1}{k}\right)^m \mathsf{E}(W)^2.$$

This implies $\mathsf{Var}(W) \le (\varepsilon^2/32)\mathsf{E}(W)^2$ and from this the lemma follows.

There are two technical issues to deal with. The first is that the distribution of each X_i^j is not exactly the target but close to it. The second is that the X_i^j's (and therefore the Y_i^j's) are independent for a fixed i but not for different i. Both issues can be handled in a manner identical to the proof of Lemma 4.2 in [25], the first using a coupling trick and the second by bounding the dependence. □

Lemma 4.3 *For $i = 0, \ldots, m$, let π_i be the distribution with density proportional to $f(x)^{a_i}$. Then $d_2(\pi_{i-1}, \pi_i) \le e$ and $d_2(\pi_i, \pi_{i-1}) \le 4$.*

Proof.

$$\begin{aligned}&d_2(\pi_{i-1}, \pi_i) \\ &= \int \frac{f(x)^{a_{i-1}} / \int f(x)^{a_{i-1}} \, dx}{f(x)^{a_i} / \int f(x)^{a_i} \, dx} \frac{f(x)^{a_{i-1}}}{\int_f (x)^{a_{i-1}} \, dx} \, dx \\ &= \frac{\int f(x)^{2a_{i-1} - a_i} \, dx \int f(x)^{a_i} \, dx}{\left(\int f(x)^{a_{i-1}} \, dx\right)^2} \\ &= \frac{Z(2a_{i-1} - a_i)Z(a_i)}{Z(a_{i-1})^2} \le \left(\frac{a_{i-1}^2}{(2a_{i-1} - a_i)a_i} \right)^n \le 4.\end{aligned}$$

On the other hand, as in the proof of Lemma 4.2,

$$d_2(\pi_i, \pi_{i-1}) = \frac{Z(2a_i - a_{i-1})Z(a_{i-1})}{Z(a_i)^2} \le e$$

□

Lemma 4.4 *For $i = 0, \ldots, m$, the distribution π_i corresponding to f_i is $O(\ln(1/\varepsilon))$-round.*

Proof. By assumption, $f_m(x) = f(x)$ is well-rounded and contained in a ball of radius $O(\sqrt{n}\ln(1/\varepsilon))$. Thus, each f_i is contained in this ball and so it follows that $\mathsf{E}(|x|^2) = O(n\ln^2(1/\varepsilon))$ for the distribution corresponding to each f_i.

Now consider the level set $L = L_f(t)$ of measure $1/8$. By assumption, this contains a ball of radius equal to some constant c. As we move through the sequence of functions, $f_m, f_{m-1}, \ldots, f_1$, the measure of L is decreasing and so for each of them, L is contained in the level set of measure $1/8$ which means this level set contains a ball of radius c. □

5 Optimization

In this section, we give an algorithm for finding the maximum of a logconcave function f, given access to f by means of an oracle that evaluates f at any desired point x and a starting point x_0 with $f(x_0) \ge \beta^n \max f$. The algorithm can be viewed as simulated annealing in a continuous setting and is suggested in [16] and analyzed for the special case when $f(x) = e^{-a^T x}$ for some vector a. Here we consider the problem of maximizing an arbitrary logconcave function. This is equivalent to finding the minimum of a convex function $g(x)$ by considering the logconcave function $e^{-g(x)}$.

We restrict the support of the function f to $K = \{y | f(y) \ge f(x_0)/2\}$. In the description below, B is an upper bound on $\ln(\max f / \min f)$. It will suffice to use $B = n\ln(2/\beta)$ from our assumption that $f(x_0) \ge \beta^n \max f$.

Optimization algorithm:

O1. Set $m = \lceil \sqrt{n} \ln \frac{2B(n+\ln(1/\delta))}{\varepsilon} \rceil$, $k = \lceil C_0 n \ln^5 n \rceil$ and for $i = 1, \ldots, m$, let

$$a_i = \frac{1}{B}(1 + \frac{1}{\sqrt{n}})^i \quad \text{and} \quad f_i(x) = f(x)^{a_i}.$$

Also $X_0^1, \ldots, X_0^k$ are independent uniform random points from K and T_0 is their covariance matrix.

O2. For $i = 1, \ldots, m$, do the following:

— Get k independent random samples $X_i^1, \ldots, X_i^k$ from π_{f_i} using hit-and-run with covariance matrix T_i and starting points $X_{i-1}^1, \ldots, X_{i-1}^k$ respectively.

— Set T_{i+1} to be the covariance matrix of $X_i^1, \ldots, X_i^k$.

O3. Output $\max_j f(X_m^j)$ and a point X_m^j that achieves the maximum.

We generate uniform random points from K, required in the first step (O1), as in [25], by putting K in near-isotropic position.

It is clear that the number of phases of the algorithm is $O^*(\sqrt{n})$. To prove Theorem 1.4, we need to show that (i) the last distribution is concentrated near the maximum, (ii) samples from the i'th phase provide a good start for the next phase, i.e., $d_2(\pi_i, \pi_i)$ is bounded and (iii) the transformation T_i keeps f_i near-isotropic and thus the time per sample remains $O^*(n^3)$.

Of these, (ii) is already given by Lemma 4.3. The next lemma establishes (i).

Lemma 5.1 *For any $\varepsilon > 0$, with*

$$m \ge \sqrt{n} \ln \frac{2B(n + \ln(1/\delta)}{\varepsilon},$$

for a random point X drawn with density proportional to f_m, we have

$$\mathsf{P}(f(X) < (1-\varepsilon)\max f) \le \delta \left(\frac{2}{e}\right)^{n-1}.$$

Proof. Let X be a random point from π_{f_m}. We can bound the desired probability as follows:

$$\begin{aligned}
&\mathsf{P}(f(X) < (1-\varepsilon)\max f) \\
&\quad = \mathsf{P}(f(X)^{a_m} < (1-\varepsilon)^{a_m}(\max f)^{a_m}) \\
&\quad = \mathsf{P}(f_m(X) < (1-\varepsilon)^{a_m} \max f_m) \\
&\quad \le \mathsf{P}(f_m(X) < e^{-2(n+\ln(1/\delta)} \max f_m).
\end{aligned}$$

Now we use Lemma 5.16 from [24] which says that for $\beta \ge 2$ (since f_m is logconcave),

$$\mathsf{P}(f_m(X) < e^{-\beta(n-1)} \max f_m) \le (e^{1-\beta}\beta)^{n-1}$$

to get the bound. □

Finally, the near-isotropy is derived using the following lemma from [16].

Lemma 5.2 ([16]) *Let μ and ν be logconcave densities with centroids z_μ and z_ν, respectively. Then for any $c \in \mathbb{R}^n$,*

$$\mathsf{E}_\mu((c \cdot (x - z_\mu))^2) \leq 16 d_2(\mu, \nu) \mathsf{E}_\nu((c \cdot (x - z_\nu))^2)$$

We apply this to $\mu = \mu_{i+1}$ and $\nu = \mu_i$ after the transformation T_i. By Lemma 2.2, the transformation T_i puts μ_i in 2-isotropic position with high probability. Using this and the second inequality of Lemma 4.3, we get that for any $c \in \mathbb{R}^n$,

$$\mathsf{E}_{\mu_{i+1}}((c \cdot (x - z_{\mu_{i+1}}))^2) \leq 128.$$

Reversing the roles of μ and ν in Lemma 2.2 gives a lower bound and proves that μ_{i+1} is 128-isotropic. Thus, the complexity of sampling is $O(n^3 \ln^5 n)$ (it suffices to set the desired variation distance to the target distribution to be $1/\text{poly}(n)$).

6 Rounding

The optimization algorithm of the previous section can be used to achieve near-isotropic position, by starting at the uniform density and stopping with $f_m = f$. The transformation T_{m+1} will put f in 2-isotropic position with high probability. The algorithm uses only a function oracle but requires $O^*(n^{4.5})$ steps. Here we give an algorithm that requires only $O^*(n^4)$ steps, but assuming we are given a point x^* where f is maximized (we will henceforth assume that $x^* = 0$).

Theorem 6.1 *Any logconcave function f can be put in near-isotropic position using $O^*(n^4)$ oracle calls, given a point that maximizes f.*

The algorithm is similar to that for a convex body given in [25]. The main difference is that instead of a *pencil* we use a *log-pencil* which is the function $G : \mathbb{R}^{n+1} \to \mathbb{R}_+$ defined as

$$G(x_0, x) = \begin{cases} f(x) & \text{if } 0 \leq x_0 \leq 2B \text{ and} \\ & \ln f(x) \geq \ln \max f - x_0, \\ 0 & \text{otherwise.} \end{cases}$$

Here B is an upper bound on $\ln(\max f / \min f)$. In words, the cross-section of G at x_0 is the level set $L_f(\max f / e^{x_0})$. It is easy to see G is also logconcave.

For $i = 1, \ldots, m$, define G_i as the restriction of G to the halfspace $\{(x, x_0) : x_0 \leq 2^{i/B}\}$ and $G_m = G$. The algorithm makes the G_i's isotropic successively. In most phases it only needs $O^*(1)$ new samples and once every n phases it uses $O^*(n)$ new samples. Once G is near-isotropic, one more round of samples can be used to make f near-isotropic in $\mathbb{R}^n$.

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. Füredi: Computing the Volume is Difficult. Discrete & Computational Geometry 2 (1987), 319-326.

[3] D. Bertsimas and S. Vempala: Solving convex programs by random walks. *J. ACM* 51(4), (2004), 540–556.

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

[5] M. E. Dyer and A. M. Frieze: Computing the volume of a convex body: a case where randomness provably helps, *Proc. of 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] A. Frieze and R. Kannan: Log-Sobolev inequalities and sampling from log-concave distributions, *Annals of Applied Probability*, **9** (1999), 14–26.

[9] A. Frieze, R. Kannan and N. Polson: Sampling from log-concave distributions, *Ann. Appl. Probab.* **4** (1994), 812–837; Correction: Sampling from log-concave distributions, ibid. **4** (1994), 1255.

[10] M. Grötschel, L. Lovász, and A. Schrijver: *Geometric Algorithms and Combinatorial Optimization*, Springer, 1988.

[11] M. Jerrum and A. Sinclair: Approximating the permanent, *SIAM Journal on Computing*, **18** (1989), 1149–1178.

[12] R. Kannan and L. Lovász: Faster mixing through average conductance, *Proc. of STOC*, 1999, 282-287.

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

[14] 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.

[15] R. Kannan, J. Mount and S. Tayur: A randomized algorithm to optimize over certain convex sets, *Math. of OR*, **20**(3), 1995, 529–549.

[16] A. Kalai and S. Vempala: Simulated Annealing for Convex Optimization. To appear in *Math of OR*.

[17] I. Koutis: On the Hardness of Approximate Multivariate Integration. *Proc. of APPROX*, (2003), 122-128.

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

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

[20] L. Lovász: Hit-and-run mixes fast, *Math. Prog.* 86 (1998), 443-461.

[21] 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.

[22] 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.

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

[24] L. Lovász and S. Vempala: The geometry of logconcave functions and sampling algorithms, to appear in *Random Structures and Algorithms*. Prelim. version in FOCS 2003. Available at: http://www-math.mit.edu/˜vempala/papers/logcon.pdf

[25] L. Lovász and S. Vempala: Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *J. Comp. Sys. Sci.* **72**(2), (2006), 392–417. Prelim. version in FOCS 2003.

[26] L. Lovász and S. Vempala: Hit-and-run from a corner. *SIAM J. Computing*, **35**(4) (2006), 985–1005. Prelim. version in STOC 2004.

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

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

[29] R.L. Smith: Efficient Monte-Carlo procedures for generating points uniformly distributed over bounded regions, *Operations Res.* **32**, (1984), 1296–1308.

[30] P. M. Vaidya: A new algorithm for minimizing convex functions over convex sets, *Mathematical Programming*, 73, 291–341, 1996.