The Geometry of Logconcave Functions and an $O^*(n^3)$ Sampling Algorithm [1]

László Lovász[2] Santosh Vempala[3]

January 2003

Technical Report
MSR-TR-2003-04

The class of *logconcave* functions in $\mathbb{R}^n$ is a common generalization of Gaussians and of indicator functions of convex sets. Motivated by the problem of sampling from a logconcave density function, we study their geometry and introduce a technique for "smoothing" them out. This leads to an efficient sampling algorithm (by a random walk) with no assumptions on the local smoothness of the density function. After appropriate preprocessing, the algorithm produces a point from approximately the right distribution in time $O^*(n^4)$, and in amortized time $O^*(n^3)$ if many sample points are needed (where the asterisk indicates that dependence on the error parameter and factors of $\log n$ are not shown).

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

[1]Part of this work was done while the second author was visiting Microsoft Research in Redmond, WA, and when the first author was visiting the Newton Institute in Cambridge, UK.

[2]Microsoft Research, One Microsoft Way, Redmond, WA 98052

[3]Department of Mathematics, Massachusetts Institute of Technology, Cambridge,

1 Introduction

Virtually all known algorithms to sample from a high dimensional convex body K (i.e., to generate a uniformly distributed random point in K) work by defining a Markov chain whose states are the points of K (or a sufficiently dense subset of it), and whose stationary distribution is uniform. Running the chain long enough produces an approximately uniformly distributed random point. The most thoroughly analyzed versions are the lattice walk [4, 7], the ball walk [12, 10, 9] and the hit-and-run walk [21, 22, 5, 13].

For what distributions is the random walk method efficient? These sampling algorithms can be extended to any other (reasonable) distribution in $\mathbb{R}^n$, but the methods for estimating their mixing time all depend on convexity properties. A natural class generalizing uniform distributions on convex sets is the class of *logconcave distributions*. For our purposes, it suffices to define these as probability distributions on the Borel sets of $\mathbb{R}^n$ which have a density function f and the logarithm of f is concave. Such density functions play an important role in stochastic optimization [19] and other applications [8]. We assume that the function is given by an *oracle*, i.e., by a subroutine that returns the value of the function at any point x. We measure the complexity of the algorithm by the number of oracle calls.

The analyses of the lattice walk and ball walk have been extended to logconcave distributions [1, 6], but these analyses need explicit assumptions on the Lipschitz constant of the distribution. In this paper, we avoid such assumptions by considering a smoother version of the function in the analysis. The smoother version is bounded by the original, continues to be logconcave, and has almost the same integral, i.e. it is almost equal at most points.

Our main result is that after appropriate preprocessing (bringing the distribution into isotropic position), the ball walk (with a Metropolis filter) can be used to generate a sample using $O^*(n^4)$ oracle calls. We get a better bound of $O^*(n^3)$ if we consider a *warm start*. This means that we start the walk not from a given point but from a random point that is already quite well distributed in the sense that its density function is at most a constant factor larger than the target density f. While this sounds quite restrictive, it is often the case (for example, when generating many sample points, or using a "bootstrapping" scheme as in [10]) that this bound gives the actual cost of the algorithm per random point. We also give an $O^*(n^5)$ algorithm for bringing an arbitrary logconcave distribution to isotropic position. Our amortized bound for sampling logconcave functions matches the best-known bound for the special case of sampling uniformly from a convex set.

Our analysis uses various geometric properties of logconcave functions; some of these are new while others are well-known or folklore, but since a reference is not readily available, we prove them in section 3.

A sequel to this paper [15] analyzes the *hit-and-run* random walk.

MA 02139. Supported by NSF Career award CCR-9875024.

2 Results

2.1 Preliminaries.

A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *logconcave* if it satisfies

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

for every $x, y \in \mathbb{R}^n$ and $0 \leq \alpha \leq 1$. This is equivalent to saying that the support K of f is convex and $\log f$ is concave on K.

An integrable function $f : \mathbb{R}^n \to \mathbb{R}_+$ is a *density function*, if $\int_{\mathbb{R}^n} f(x)\,dx = 1$. Every non-negative integrable function f gives rise to a probability measure on the measurable subsets of $\mathbb{R}^n$ defined by

$$\pi_f(S) = \int_S f(x)\,dx \Big/ \int_{\mathbb{R}^n} f(x)\,dx\ .$$

The *centroid* of a density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is the point

$$z_f = \int_{\mathbb{R}^n} f(x)x\,dx;$$

the *covariance matrix* of the function f is the matrix

$$V_f = \int_{\mathbb{R}^n} f(x)xx^T\,dx$$

(we assume that these integrals exist).

For any logconcave function $f :\ \mathbb{R} \to \mathbb{R}^n$, we denote by M_f its maximum value. We denote by

$$L_f(t) = \{x \in \mathbb{R}^n :\ f(x) \geq t\}$$

its level sets, and by

$$f_t(x) = \begin{cases} f(x), & \text{if } f(x) \geq t, \\ 0, & \text{if } f(x) < t, \end{cases}$$

its restriction to the level set. It is easy to see that f_t is logconcave. In M_f and L_f we omit the subscript if f is understood.

A density function $f : \mathbb{R}^n \to \mathbb{R}_+$ is *isotropic*, if its centroid is 0, and its covariance matrix is the identity matrix. This latter condition can be expressed in terms of the coordinate functions as

$$\int_{\mathbb{R}^n} x_i x_j f(x)\,dx = \delta_{ij}$$

for all $1 \leq i, j \leq n$. This condition is equivalent to saying that for every vector $v \in \mathbb{R}^n$,

$$\int_{\mathbb{R}^n} (v \cdot x)^2 f(x)\,dx = |v|^2.$$

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

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

We say that f is *near-isotropic up to a factor of* C, if $(1/C) \le \int (u^T x)^2 \, d\pi_f(x) \le C$ for every unit vector u. We extend the notions of "isotropic" and "non-isotropic" to non-negative integrable functions f, in which case we mean that the density function $f / \int_{\mathbb{R}^n} f$ is isotropic or near-isotropic.

Given any density function f with finite second moment $\int_{\mathbb{R}^n} \|x\|^2 f(x)\, dx$, there is an affine transformation of the space bringing it to isotropic position, and this transformation is unique up to an orthogonal transformation of the space.

For two points $u, v \in \mathbb{R}^n$, we denote by $d(u, v)$ their euclidean distance. For two probability distributions σ, τ on the same underlying σ-algebra, let

$$d_{\text{tv}}(\sigma, \tau) = \sup_A (\sigma(A) - \tau(A))$$

be their total variation distance.

2.2 The random walk.

Let f be a logconcave distribution in $\mathbb{R}^n$.

The ball walk (with a Metropolis filter) to sample from f is defined as follows:

- Pick a uniformly distributed random point y in the ball of radius r centered at the current point.
- Move to y with probability $\min(1, f(y)/f(x))$; stay at the current point with the remaining probability.

In connection with the second step, we have to discuss how the function is given: we assume that it is given by an *oracle*. This means that for any $x \in \mathbb{R}^n$, the oracle returns the value $f(x)$. (We ignore here the issue that if the value of the function is irrational, the oracle only returns an approximation of f.) It would be enough to have an oracle which returns the value $C \cdot f(x)$ for some unknown constant $C > 0$ (this situation occurs in many sampling problems e.g. in statistical mechanics and simulated annealing).

For technical reasons, we also need a "guarantee" from the oracle that the centroid z_f of f satisfies $\|z_f\| \le Z$ and that all the eigenvalues of the covariance matrix are between r and R, where Z, r and R are given positive numbers.

Our main theorem concerns functions that are near-isotropic (up to some fixed constant factor c). In section 2.4, we discuss how to preprocess the function in order to achieve this.

Theorem 2.1 *If f is near-isotropic, then it can be approximately sampled in time $O^*(n^4)$ and in amortized time $O^*(n^3)$ if more than n sample points are needed; any logconcave function can be brought into isotropic position in time $O^*(n^5)$.*

Theorem 2.1 is based on the following more explicit result about a "warm start".

Theorem 2.2 *Let f be a logconcave density function in $\mathbb{R}^n$ that is near-isotropic up to a factor of c. Let σ be a starting distribution and let σ^m be the distribution of the current point after m steps of the ball walk. Assume that there is a $D > 0$ such that $\sigma(S) \leq D\pi_f(S)$ for every set S. Then for*

$$m > 10^{10} c^2 D^2 \frac{n^3}{\varepsilon^2} \log \frac{1}{\varepsilon},$$

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

2.3 Isoperimetry.

As in all papers since [4], bounding the mixing time depends on a geometric isoperimetric inequality.

We will use an extension of Theorem 5.2 from [10] to logconcave functions. For a logconcave density function f, define

$$M(f) = \mathsf{E}_f(|x - z_f|).$$

Theorem 2.3 *Let f be a logconcave density function on $\mathbb{R}^n$. For any partition of $\mathbb{R}^n$ into three measurable sets S_1, S_2, S_3,*

$$\pi_f(S_3) \geq \frac{\ln 2}{M(f)} d(S_1, S_2)\pi_f(S_1)\pi_f(S_2).$$

Note that for an isotropic density function in $\mathbb{R}^n$, $M(f) \leq \sqrt{n}$.

2.4 Transforming to isotropic position.

The results of section 2.2 above use the assumption that the given density function is near-isotropic. To bring the function to this position can be considered as a preprocessing step, which only needs to be done once, and then we can generate any number of independent samples at the cost of $O^*(n^3)$ or $O^*(n^4)$ oracle calls per sample, as described earlier. But the problem of transforming into near-isotropic position is closely intertwined with the sampling problem, since we use sampling to transform an arbitrary logconcave density function into near-isotropic position.

In this section we describe an algorithm to achieve near-isotropic position. The following theorem is a consequence of a generalized Khinchine's Inequality (Theorem 3.17) and Rudelson's theorem [20] and is the basis of the algorithm.

Theorem 2.4 *Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an isotropic logconcave function. Let $v_1, v_2, \ldots, v_m$ be independent samples from π_f with*

$$m = 100 \cdot \frac{n}{\eta^2} \cdot \log^3 \frac{n}{\eta^2}.$$

Then

$$E\Big\|\frac{1}{m}\sum_{i=1}^{m} v_i v_i^T - I\Big\| \le \eta.$$

A corollary of this result is the following:

Corollary 2.5 *Let $f: \mathbb{R}^n \to \mathbb{R}_+$ be a (not necessarily isotropic) logconcave function. Let $v_1, v_2, \ldots, v_m$ be independent samples from π_f (where m is defined as above). Compute the matrices $V = \frac{1}{m}\sum_{i=1}^m v_i v_i^T$ and $W = V^{-1/2}$. Then with probability at least $1-(\eta/\delta)$, we have*

$$\|V - V_f\| < \delta,$$

and hence the transformed function $\hat{f}(x) = f(Wx)$ is near-isotropic up to a factor of $1/(1-\delta)$.

The error probability of η/δ is not good enough for us; we cannot choose η small because the number of sample points m depends badly on η. The following trick reduces the error probability for a moderate cost:

Algorithm A.

- Choose $\eta = 1/100$ and $\delta = 1/10$.
- Repeat the construction in Corollary 2.5 q times, to get q matrices $V^{(1)}, \ldots, V^{(q)}$.
- For each of these matrices $V^{(i)}$, count how many other matrices $V^{(j)}$ satisfy

  $$\|V^{(j)} - V^{(i)}\|) < .2. \tag{1}$$

  If you find one for which this number is larger than $q/2$, return this as an approximation of V_f. Otherwise, the procedure fails.

Theorem 2.6 *With probability at least $1-(4/5)^q$, Algorithm A returns a matrix V satisfying $\|V - V_f\| < .3$.*

Proof. For each $1 \le j \le k$, we have

$$\mathsf{P}\Big(\|V^{(j)} - V_f\|) < .1\Big) > .9. \tag{2}$$

Hence (by Chernoff's inequality) with probability at least $1-(4/5)^k$, more than half of the $V^{(j)}$ satisfy (2). If both $V^{(i)}$ and $V^{(j)}$ satisfy (2), then they clearly satisfy (1), and hence in this case algorithm cannot fail. Furthermore, if $V^{(i)}$ is returned, then by pigeon hole, there is a j such that both (2) and (1) are satisfied, and hence $\|V^{(i)} - V_f\| < .3$ as claimed. □

Now the algorithm to bring f to near-isotropic position can be sketched as follows. Define

$$T_k = \frac{M_f}{2^{(1+1/n)^k}} \qquad k = 0, 1, \ldots$$

The algorithm is iterative and in the k-th phase it brings the function $g_k = f_{T_k}$ into near-isotropic position.

Algorithm B.

- Choose $p = Cn \log n$ and $q = \log(p/\varepsilon)$.
- Bring the level set $\{x : f(x) \geq M/2\}$ to near-isotropic position (using e.g. the algorithm from [10]).
- For $k = 0, 1, \ldots, p$, compute an approximation V of V_{g_k} using Algorithm A, and apply the linear transformation $V^{-1/2}$.

By Corollary 3.15, we have that after phase k, not only g_k is approximately isotropic, but also g_{k+1} is near-isotropic; in addition a random sample from g_k provides a warm start for sampling from g_{k+1} in the next phase. So we need to walk only $O^*(n^3)$ steps to get a point from (approximately) g_{k+1}. By Lemma 3.11 together with Theorem 3.9(e), the measure of the function outside the set $\{x : f(x) \geq M/T_{Cn \log n}\}$ is negligible and so $Cn \log n$ phases suffice to bring f itself to near-isotropic position.

The above description of the algorithm suits analysis, but is not how one implements it. We don't transform the body, rather, we transform the euclidean norm of the space. More precisely, we maintain a basis $u_1, \ldots, u_n$, which is initialized as $u_i = e_i$ in phase 0. The u_i are fixed throughout each phase. To generate a line through the current point, we generate n independent random numbers $X_1, \ldots, X_n$ from the standard Gaussian distribution, and compute the vector $X_1 u_1 + \ldots X_n u_n$. The line we move on will be the line containing this vector.

At the end of phase k, we have generated m independent random points $v_1, \ldots, v_m$ from the distribution $f_{T_{k+1}}$. We compute the matrix $W = \left(\frac{1}{m} \sum_{i=1}^m v_i v_i^T\right)^{1/2}$ and update the vectors u_i by letting $u_i = W e_i$.

Each sample (after the first) takes $O^*(n^3)$ oracle calls and hence each phase takes $O^*(n^5)$ calls. The overall complexity of the algorithm is $O^*(n^5)$ oracle calls.

3 The geometry of logconcave functions.

In this section we state geometric properties of logconcave functions that are used in the analysis of our algorithm. Most of these facts are well-known or even folklore, but references are not easy to pin down. Since the constants involved are needed to formalize our algorithm (not only to analyze it), we found that we have to include this section containing the proofs with explicit constants.

We start with some definitions. The *marginals* of a function $f : \mathbb{R}^n \to \mathbb{R}_+$ are defined by

$$G(x_1, \ldots, x_k) = \int_{\mathbb{R}^{n-k}} f(x_1, \ldots, x_n)\, dx_{k+1} \ldots dx_n.$$

The *first marginal*

$$g(t) = \int_{x_2,\dots,x_n} f(t, x_2, \dots, x_n)\, dx_2 \dots dx_n$$

will be used most often. It is easy to check that if f is in isotropic position, then so are its marginals. The *distribution function* of f is defined by

$$F(t_1, \dots, t_n) = \int_{x_1 \le t_1, \dots, x_n \le t_n} f(x_1, \dots, x_n)\, dx_1 \dots dx_n.$$

Clearly, the product and the minimum of logconcave functions is logconcave. The sum of logconcave functions is not logconcave in general; but the following fundamental properties of logconcave functions, proved by Dinghas [3], Leindler [16] and Prékopa [17, 18], can make up for this in many cases.

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.*

3.1 One-dimensional functions

Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be an integrable function such that $g(x)$ tends to 0 faster than any polynomial as $x \to \infty$. Define its moments, as usual, by

$$M_g(n) = \int_0^\infty t^n g(t)\, dt.$$

Lemma 3.2 (a) *The sequence* $(M_g(n) : \ n = 0, 1, \dots)$ *is logconvex.*

(b) *If g is monotone decreasing, then the sequence defined by*

$$M'_g(n) = \begin{cases} nM_g(n-1)), & \text{if } n > 0, \\ g(0), & \text{if } n = 0 \end{cases}$$

is also logconvex.

(c) *If g is logconcave, then the sequence $M_g(n)/n!$ is logconcave.*

(d) *If g is logconcave, then*

$$g(0)M_g(1) \le M_g(0)^2.$$

(i.e., we could append $g(0)$ at the beginning of the sequence in (c) and maintain logconcavity).

Proof. (a) We have for every real x

$$0 \le \int_0^\infty (t+x)^2 t^n g(t)\, dt = x^2 M_g(n) + 2xM_g(n+1) + M_g(n+2).$$

Hence the discriminant of the quadratic polynomial on the right hand side must be non-positive:

$$M_g(n+1)^2 - M_g(n)M_g(n+2) \le 0,$$

which is just the logconvexity of the sequence.

(b) We may assume (by approximating) that g is differentiable. Then $-g'$ is nonnegative, which by (a) implies that the sequence $M_{-g'}(n)$ is logconvex. Integrating by parts,

$$M_{-g'}(n) = -\int_0^\infty t^n g'(t)\,dt = -[t^n g(t)]_0^\infty + \int_0^\infty nt^{n-1} g(t)\,dt = nM_g(n-1)$$

for $n \geq 1$, and

$$M_{-g'}(0) = -\int_0^\infty g'(t)\,dt = g(0).$$

This proves (b).

(c) Let $h(t) = \beta e^{-\gamma t}$ be an exponential function ($\beta, \gamma > 0$) such that

$$M_h(n) = M_g(n) \qquad \text{and} \qquad M_h(n+2) = M_g(n+2)$$

(it is easy to see that such an exponential function exists). Then we have

$$\int_0^\infty t^n (h(t) - g(t))\,dt = 0, \qquad \int_0^\infty t^{n+2} (h(t) - g(t))\,dt = 0.$$

From this it follows that the graph of h must intersect the graph of g at least twice. By the logconcavity of g, the graphs intersect at exactly two points $a < b$. Furthermore, $h \leq g$ in the interval $[a, b]$ and $h \geq g$ outside this interval. So the quadratic polynomial $(t-a)(t-b)$ has the same sign pattern as $h - g$, and hence

$$\int_0^\infty (t-a)(t-b) t^n (h(t) - g(t))\,dt \geq 0.$$

Expanding and rearranging, we get

$$\int_0^\infty t^{n+2}(h(t) - g(t))\,dt + ab \int_0^\infty t^n (h(t) - g(t))\,dt \geq (a+b) \int_0^\infty t^{n+1} (h(t) - g(t))\,dt.$$

The left hand side is 0 by the definition of h, so the right hand side is nonpositive, which implies that

$$M_h(n+1) = \int_0^\infty t^{n+1} h(t)\,dt \leq \int_0^\infty t^{n+1} g(t)\,dt = M_g(n+1).$$

In other words, this shows that it is enough to verify the inequality for exponential functions. But for these functions, the inequality holds trivially.

(d) The proof is similar and is left to the reader. □

Lemma 3.3 *Let X be a random point drawn from a one-dimensional logconcave distribution. Then*

$$\mathsf{P}(X \geq \mathsf{E}X) \geq \frac{1}{e}.$$

Proof. We may assume without loss of generality that $\mathsf{E}X = 0$. It will be convenient to assume that $|X| \le K$; the general case then follows by approximating a general logconcave distribution by such distributions.

Let $G(x) = \mathsf{P}(X \le x)$. Then G is logconcave, monotone increasing, and we have $G(x) = 0$ for $x \le -K$ and $G(x) = 1$ for $x \ge K$. The assumption that z is the centroid implies that

$$\int_{-K}^{K} xG'(x)\,dx = 0,$$

which by partial integration means that

$$\int_{-K}^{K} G(x)\,dx = K.$$

We want to prove that $G(0) \ge 1/e$.

The function $\ln G$ is concave, so it lies below its tangent at 0; this means that $G(x) \le G(0)e^{cx}$, where $c = G'(0)/G(0) > 0$. We may choose K large enough so that $1/c < K$. Then

$$G(x) \le \begin{cases} G(0)e^{cx} & \text{if } x \le 1/c, \\ 1 & \text{if } x > 1/c, \end{cases}$$

and so

$$\begin{aligned} K &= \int_{-K}^{K} G(x),\, dx \\ &\le \int_{-\infty}^{1/c} G(0)e^{cx}\,dx + \int_{1/c}^{K} 1\,dx \\ &= \frac{eG(0)}{c} + K - \frac{1}{c}, \end{aligned}$$

which implies that $G(0) \ge 1/e$ as claimed. $\square$

Lemma 3.4 *Let $g : \mathbb{R} \to \mathbb{R}_+$ be an isotropic logconcave density function.*
(a) *For all x, $g(x) \le 1$.*
(b) $g(0) \ge \frac{1}{8}$.

Proof. (a) Let the maximum of g be attained at a point z, and suppose that $g(z) > 1$. For $i = 0, 1, 2$, let

$$M_i = \int_{z}^{\infty} (x - z)^i\, g(x)\,dx,$$

$$N_i = \int_{-\infty}^{z} (z - x)^i\, g(x)\,dx.$$

Clearly

$$M_0 + N_0 = 1, \qquad N_1 - M_1 = z, \qquad M_2 + N_2 = 1 + z^2.$$

So

$$\begin{aligned}
M_2 + N_2 &= (M_0 + N_0)^2 + (M_1 - N_1)^2 \\
&= (M_0 - M_1)^2 + (N_0 - N_1)^2 + 2(M_0N_0 - M_1N_1) + 2(M_0M_1 + N_0N_1) \\
&\geq 2(M_0M_1 + N_0N_1),
\end{aligned}$$

since by Lemma 3.2(d), we have $M_1 \leq M_0^2/g(z) \leq M_0$, and similarly $N_1 \leq N_0^2/g(z) \leq N_0$.

On the other hand, by Lemma 3.2(c) and (d), we have

$$M_2 \leq 2M_1^2/M_0 \leq 2M_1M_0/g(z) < 2M_1M_0, \qquad N_2 < 2N_1N_0,$$

and so

$$M_2 + N_2 < 2(M_0M_1 + N_0N_1),$$

a contradiction proving (a).

(b) We may assume that $g(x)$ is monotone decreasing for $x \geq 0$. Let g_0 be the restriction of g to the nonnegative semiline. By Lemma 3.2(b),

$$M_{g_0}(0) \leq g(0)^{2/3}(3M_{g_0}(2))^{1/3}.$$

Here trivially $M_{g_0}(2) \leq M_g(2) = 1$, while Lemma 3.3 implies that $M_{g_0}(0) \geq (1/e)M_g(0) = 1/e$. Substituting these bounds, we get

$$g(0) \geq \sqrt{\frac{1}{3e^3}} \geq \frac{1}{8}.$$

□

Lemma 3.4(a) is tight, as shown by the function

$$g(x) = \begin{cases} e^{-1-x}, & \text{if } x \geq -1, \\ 0, & \text{if } x < -1. \end{cases}$$

Part (b) is not tight; most probably the right constant is $1/(2 \cdot \sqrt{3})$, attained by the uniform distribution on the interval $[-\sqrt{3}, \sqrt{3}]$.

Lemma 3.5 *Let X be a random point drawn from a logconcave density function $g : \mathbb{R} \to \mathbb{R}_+$.*

(a) *For every $c \geq 0$,*

$$\mathsf{P}(g(X) \leq c) \leq \frac{c}{M_g}.$$

(b) *For every $0 \leq c \leq g(0)$,*

$$\mathsf{P}(\min g(2X), g(-2X) \leq c) \geq \frac{c}{4g(0)}.$$

Proof. (a) We may assume that the maximum of g is assumed at 0. Let $q > 0$ be defined by $g(q) = c$, and let $h(t) = g(0)e^{-\gamma t}$ be an exponential function such that $h(q) = g(q)$. Clearly such a γ exists, and $h(0) = g(0)$, $\gamma > 0$. By the logconcavity of the function, the graph of h is below the graph of g between 0 and q, and above outside. Hence

$$\frac{\int_q^\infty g(t)\,dt}{\int_0^\infty g(t),\,dt} \le \frac{\int_q^\infty h(t)\,dt}{\int_0^\infty h(t),\,dt}$$

Here

$$\int_0^\infty h(t)\,dt = \frac{g(0)}{\gamma}, \int_q^\infty h(t)\,dt = \frac{g(q)}{\gamma} = \frac{c}{\gamma},$$

and so we get that

$$\int_q^\infty g(t)\,dt \le \frac{c}{g(0)} \int_0^\infty g(t)\,dt.$$

Here $g(0) = M_g$, furthermore

$$\int_q^\infty g(t)\,dt = \mathsf{P}(X > q) = \mathsf{P}(g(X) < c, X > 0),$$

and

$$\int_0^\infty g(t)\,dt = \mathsf{P}(X > 0),$$

so we get that

$$\mathsf{P}(g(X) < c, X > 0) \le \frac{c}{M_g}\mathsf{P}(X > 0).$$

Similarly,

$$\mathsf{P}(g(X) < c, X < 0) \le \frac{c}{M_g}\mathsf{P}(X < 0).$$

Adding up these two inequalities, the assertion follows.

(b) We may assume that $\mathsf{P}(X \le 0) \ge 1/2$. Let $q > 0$ be defined by $g(q) = c$. If $\mathsf{P}(X \le q/2) > 1/4$ then the conclusion is obvious, so suppose that $\mathsf{P}(X \ge q/2) \le 1/4$. Similarly, we can assume that $\mathsf{P}(-q/2 < X < 0) \ge \frac{1}{4}$.

Let $h(t) = \beta e^{-\gamma t}$ be an exponential function such that

$$\int_{-q/2}^0 h(t)\,dt = \int_{-q/2}^0 g(t)\,dt = a,$$

and

$$\int_{q/2}^q h(t)\,dt = \int_{q/2}^q g(t)\,dt = b.$$

It is easy to see that such β and γ exist, and that $\beta, \gamma > 0$. From the definition of h we have

$$\begin{aligned} \frac{\beta}{\gamma}(e^{\gamma q/2} - 1) &= a, \\ \frac{\beta}{\gamma}(e^{-\gamma q/2} - e^{-\gamma q} &= b. \end{aligned}$$

Dividing these equations with each other, we obtain

$$e^{\gamma q} = \frac{a}{b}.$$

By the logconcavity of the function, the graph of h must intersect the graph of g in a point in the interval $[-q/2, 0]$ as well as in a point in the interval $[q/2, q]$; it is below the graph of g between these two points and above outside. In particular, we get

$$c = g(q) \le h(q) = \beta e^{-\gamma q} = \beta \frac{b}{a},$$

and

$$g(0) \ge h(0) = \beta,$$

so

$$b \ge a \frac{c}{g(0)} \ge \frac{c}{4g(0)}.$$

□

We conclude with a useful lemma about the exponential function.

Lemma 3.6 *Let g be an exponential function restricted to an interval, i.e.,*

$$g(x) = \begin{cases} \beta e^{\gamma x}, & \text{if } a \le x \le b, \\ 0, otherwise. \end{cases}$$

such that

$$\int_{-\infty}^{\infty} (x^2 - 1) g(x)\, dx \le 0. \tag{3}$$

Then,

$$\int_{|x|\le R} g(x)\, dx \ge (1 - e^{-R-1}) \int_{-\infty}^{\infty} g(x)\, dx. \tag{4}$$

Proof. Clearly we may assume that $\gamma \ge 0$, $a \le -1$, $b \ge 1$, and either $a \le -R$ or $b \ge R$. Further elementary considerations show that we can assume that equality holds in (3). Then from $\gamma \ge 0$ it follows that $|a| \ge b$, so $a < -R$. If $b > 1$, then we can decrease a and increase b so that (3) is maintained, and (4) will still fail if it failed to begin with. So we may assume that either $b = 1$ or $a = -\infty$.

If $b = 1$, then we can write (4) as

$$\frac{e^{\gamma} - e^{-\gamma R}}{e^{\gamma} - e^{\gamma a}} \ge 1 - e^{-(R+1},$$

which is easily seen to hold if $\gamma \ge 1$, so suppose that $\gamma \le 1$. But then (3) does not hold unless with equality unless $a = -\infty$.

So assume that $a = -\infty$. In this case, (4) can be simplified to

$$\gamma(R + b) \ge R + 1. \tag{5}$$

Equality in (3) implies that $b = (1 + \sqrt{\gamma^2 - 1})/\gamma$. So $\gamma \ge 1$, which implies (5). □

3.2 Higher dimensional functions

Now consider a logconcave density function $f: \mathbb{R}^n \to \mathbb{R}$, where we assume that $n > 5$ to exclude some trivial complications.

Lemma 3.3 extends to any dimension without difficulty:

Lemma 3.7 *Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be a logconcave density function, and let H be any halfspace containing its centroid z. Then*

$$\int_H f(x)\,dx \geq \frac{1}{e}.$$

Proof. We may assume without loss of generality H is orthogonal to the first axis. Then the assertion follows by applying Lemma 3.3 to the first marginal of f. $\square$

Lemma 3.8 *Let L be a level set of an isotropic logconcave density function f on $\mathbb{R}^n$. If L does not contain a ball of radius t, then $\int_L f(x)dx \leq et$.*

Proof. Let

$$h(x) = \begin{cases} f(x), & \text{if } x \in L, \\ 0, & \text{otherwise.} \end{cases}$$

Then h is logconcave. Let z be the centroid of h. Assume that L does not contain a ball of radius t. Then by the convexity of L, there exists $u \in \mathbb{R}^n$, $|u| = 1$ such that

$$\max_{x \in L} u^T(x - z) < t.$$

Rotate the coordinates so that $u = (1, 0, \ldots, 0)^T$, i.e.

$$\max_{x \in L} x_1 - z_1 < t.$$

The first marginal g of f is also logconcave. Further since f is in isotropic position, so is g. Lemma 3.4 implies that $g \leq 1$. Hence

$$\int_{x_1 \geq z_1} h(x)\,dx \leq \int_{z_1 \leq x_1 \leq z_1 + t} f(x)\,dx$$

$$= \int_{z_1 \leq x_1 \leq z_1 + t} g(x_1)\,dx_1 \leq t.$$

This bounds the probability of one "half" of L. But since h is a logconcave function, we have by Lemma 3.3 that

$$\int_L f(x)\,dx = \int_{\mathbb{R}^n} h(x)\,dx \leq et.$$

$\square$

Theorem 3.9 *Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an isotropic logconcave density function.*

(a) *For every $v \in \mathbb{R}^n$ with $0 \le |v| \le 1/9$, we have $2^{-9n|v|} f(0) \le f(v) \le 2^{9n|v|} f(0)$.*

(b) *$f(x) \le 2^{4n} f(0)$ for every x.*

(c) *There is an $x \in \mathbb{R}^n$ such that $f(x) > 1/\sqrt{2e\pi}^n$.*

(d) *$2^{-7n} \le f(0) \le 10^n n^{n/2}$.*

(e) *$f(x) \le 2^{8n} n^{n/2}$ for every x.*

Proof. (a) We prove the lower bound; the upper bound is analogous. Suppose that there is a point u with $|u| = t \le 1/9$ and $f(u) < 2^{-9nt} f(0)$. Let $v = (1/(9t))u$, then by logconcavity, $f(v) < 2^{-n} f(0)$. Let H be a hyperplane through v supporting the convex set $\{x \in \mathbb{R}^n : f(x) \ge f(v)\}$. We may assume that H is the hyperplane $x_1 = a$ for some $0 < a \le 1/9$. So $f(x) < 2^{-n} f(0)$ for every x with $x_1 = a$.

Let g be the first marginal of f. Then g is also isotropic, and hence $g(y) \le 1$ for all y, by Lemma 3.4. We also know by Lemma 3.3 that

$$\int_0^\infty g(y)\, dy \ge \frac{1}{e}.$$

We claim that

$$g(2a) \le \frac{g(a)}{2}. \tag{6}$$

Indeed, using the logconcavity of f, we get for every x with $x_1 = a$

$$f(2x) \le \frac{f(x)^2}{f(0)} \le \frac{f(x)}{2^n},$$

and hence

$$g(2a) = \int_{(x_1=2a)} f(x)\, dx_2 \dots dx_n \le 2^{-n} 2^{n-1} \int_{(x_1=a)} f(x)\, dx_2 \dots dx_n = \frac{g(a)}{2}.$$

Inequality (6) implies, by the logconcavity of g, that $g(x+a) \le g(x)/2$ for every $x \ge a$. Hence

$$\int_a^\infty g(y)\, dy \le 2 \int_a^{2a} g(y)\, dy \le 2a,$$

and hence

$$\int_0^\infty g(y)\, dy = \int_0^a g(y)\, dy + \int_a^\infty g(y)\, dy \le 3a < \frac{1}{e},$$

a contradiction.

(b) The proof is similar to the proof of (a). Let w be an arbitrary point, and suppose that $f(w) > 2^{4n} f(0)$. Let H be a hyperplane through 0 supporting the convex set $\{x \in \mathbb{R}^n : f(x) \ge f(0)\}$. We may assume that H is the hyperplane $x_1 = 0$. So $f(x) \le f(0)$ for every x with $x_1 = 0$. Let g be the first marginal of f. Let H_t denote the hyperplane $x_1 = t$. We may assume that $w \in H_b$ with $b > 0$.

Let $a = b/2$, let x be a point on H_a, and let x' be the intersection point of the line through w and x with the hyperplane H_0. Then by logconcavity,

$$f(w)f(x) \le f(x')^2,$$

whence

$$f(x') \ge f(w)^{1/2} f(x)^{1/2} \ge \left(\frac{f(w)}{f(x)}\right)^{1/2} f(x) \ge 4^n f(x),$$

and hence

$$g(a) = \int_{H_a} f(x)\,dx \ge 2^{1-n}4^n \int_{H_0} f(x)\,dx = 2^{1-n}4^n g(0),$$

and so, using Lemma 3.4(b), we get

$$g(a) \ge \frac{2^n}{4}$$

This contradicts Lemma 3.4(a).

(c) Suppose not. For a random point X from the distribution, we have $\mathsf{E}(|X|^2) = n$, and hence by Markov's inequality, $\mathsf{P}(|X|^2 \le 2n) \ge 1/2$. In other words,

$$\int_{\sqrt{2n}B} f(x)\,dx \ge \frac{1}{2}.$$

On the other hand,

$$\int_{\sqrt{2n}B} f(x)\,dx \le 6^{-n}\mathrm{vol}(\sqrt{2n}B) = 6^{-n}(2n)^{n/2}\pi_n < \frac{1}{2},$$

which is a contradiction.

(d) The lower bound follows from parts (b) and (c). Part (a) implies that

$$\int_{\mathbb{R}^n} f(x)\,dx \ge \int_{|x|\le 1/9} f(x)\,dx \ge \pi_n \frac{1}{9^n}\frac{f(0)}{2^n},$$

and since this integral is 1, we get that

$$f(0) \le \frac{18^n}{\pi_n} < 20^n n^{n/2}.$$

(e) is immediate from (d). □

Lemma 3.10 *Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an isotropic logconcave density function. Then for every line ℓ through* 0,

$$\int_\ell f(x)\,d\lambda_\ell(x) \ge 2^{-7n}.$$

Proof. We may assume that ℓ is the x_n-axis. Consider the marginal

$$h(x_1,\dots,x_{n-1}) = \int_{-\infty}^{\infty} f(x_1,\dots,x_{n-1},t)\,dt.$$

This is also an isotropic logconcave density function, so by Theorem 3.9(d) , we have

$$\int_\ell f(x)\,d\lambda_\ell(x) = h(0) \ge 2^{-7(n-1)} > 2^{-7n}.$$

$\square$

The following lemma generalizes Lemma 3.5(a) to arbitrary dimension.

Lemma 3.11 *Let X be a random point drawn from a distribution with a logconcave density function $f:\mathbb{R}^n \to \mathbb{R}_+$. If $\beta \ge 2$, then*

$$\mathsf{P}(f(X) \le e^{-\beta(n-1)}) \le (e^{1-\beta}\beta)^{n-1}.$$

Proof. We may assume that f is continuous, M_f is attained at the origin, and $M_f = f(0) = 1$. Let $c = e^{-\beta(n-1)}$. Using polar coordinates, we have

$$1 = \int_{\mathbb{R}^n} f(x)\,dx = \int_{u\in S}\int_0^\infty f(tu)t^{n-1}\,dt$$

and

$$\mathsf{P}(f(X)\le c) = \int_{f(x)\le c} f(x)\,dx = \int_{u\in S}\int_{t:\,f(tu)\le c} f(tu)t^{n-1}\,dt$$

Fix a unit vector u, and let $q>0$ be defined by $f(qu)=c$. By logconcavity,

$$f(tu)\begin{cases} \ge e^{\beta(n-1)t/q}, & \text{if } t\le q,\\ \le e^{\beta(n-1)t/q}, & \text{if } t \ge q.\end{cases}$$

Hence

$$\frac{\int_q^\infty f(tu)t^{n-1}\,dt}{\int_0^\infty f(tu)t^{n-1}\,dt} \le \frac{\int_q^\infty e^{\beta(n-1)t/q}t^{n-1}\,dt}{\int_0^\infty e^{\beta(n-1)t/q}t^{n-1}\,dt}.$$

The integrals on the right hand side can be evaluated:

$$\int_0^\infty e^{\beta(n-1)t/q}t^{n-1}\,dt = (n-1)!\left(\frac{\beta(n-1)}{q}\right)^n,$$

and

$$\int_q^\infty e^{\beta(n-1)t/q}t^{n-1}\,dt = (n-1)!\left(\frac{\beta(n-1)}{q}\right)^n e^{-\beta(n-1)}\sum_{k=0}^{n-1}\frac{\lambda^k}{k!}.$$

The last sum can be estimated by $2(\beta(n-1))^{n-1}/(n-1)! < (e\beta)^{n-1}$. Thus

$$\int_q^\infty f(tu)t^{n-1}\,dt \le (e^{1-\beta}\beta)^{n-1}\int_0^\infty f(tu)t^{n-1}\,dt,$$

and so

$$\begin{aligned}\mathsf{P}(f(X) \le c) &= \int_{u\in S}\int_{t:\, f(tu)\le c} f(tu)t^{n-1}\,dt \\ &\le (e^{1-\beta}\beta)^{n-1}\int_{u\in S}\int_0^\infty f(tu)t^{n-1}\,dt \\ &= (e^{1-\beta}\beta)^{n-1}.\end{aligned}$$

□

Lemma 3.12 *Let $X \in \mathbb{R}^n$ be a random point from an isotropic logconcave distribution. Then for any $R > 1$, $\mathsf{P}(|X| > R) < e^{-R}$.*

Proof. Since the distribution is isotropic, we have

$$\int_{\mathbb{R}^n} (|x|^2 - 1)f(x)\,dx = 0,$$

and if the assertion is false, then

$$\int_{|x|>R} f(x)\,dx - e^{-R}\int_{\mathbb{R}^n} f(x)\,dx > 0.$$

By the Localization Lemma, we have two points a, b and a $\gamma > 0$ so that

$$\int_0^1 (|(1-t)a + tb|^2 - 1)e^{\gamma t}\,dt = 0, \tag{7}$$

and

$$\int_{\substack{0\le t\le 1\\ |a+tv|>R}} e^{\gamma t}\,dt - e^{-R}\int_0^1 e^{\gamma t}\,dt > 0. \tag{8}$$

It will be convenient to re-parametrize this segment $[a, b]$ by considering the closest point v of its line to the origin, and a unit vector u pointing in the direction of $b - a$. Let $R' = \sqrt{R^2 - |v|^2}$, then we can rewrite (7) and (8) as

$$\int_{t_1}^{t_2} (|u|^2 + t^2 - 1)e^{\gamma t}\,dt = 0, \tag{9}$$

and

$$\int_{\substack{t_1\le t\le t_2\\ |t|>R'}} e^{\gamma t}\,dt - e^{-R}\int_{t_1}^{t_2} e^{\gamma t}\,dt > 0 \tag{10}$$

(with some t_1, t_2). Equation (9) implies that $|u| < 1$. Introducing the new variable $s = t/\sqrt{1-|u|^2}$, we get

$$\int_{s_1}^{s_2} (s^2 - 1)e^{\kappa s}\,ds = 0,$$

and

$$\int_{\substack{s_1\le s\le s_2\\ |s|>R''}} e^{\kappa s}\,ds - e^{-R}\int_{s_0}^{s_1} e^{\kappa s}\,ds > 0,$$

where $R'' = R'/\sqrt{1-|u|^2}$, and s_1, s_2, κ are similarly transformed. Now this contradicts Lemma 3.6. □

The following lemma generalizes the upper bound in Theorem 4.1 of [11]:

Lemma 3.13 *Let $f:\ \mathbb{R}^n \to \mathbb{R}$ be an isotropic logconcave function, and let z be a point where it assumes its maximum. Then $|z| \le n+1$.*

The characteristic function of an isotropic regular simplex shows that the bound is essentially tight.

Proof. Write $x = x + tu$, where $t \in \mathbb{R}_+$ and $|u| = 1$. Let $e = z/|z|$. We can write

$$1 = \int_{\mathbb{R}^n} (e^{\mathsf{T}}x)^2 f(x)\,dx = \int_{|u|=1}\int_0^{\infty} (e^{\mathsf{T}}(z+tu))^2 f(tu)t^{n-1}\,dt\,du.$$

Fix any u, and let $g(t) = f(tu)$. Then the inside integral is

$$\int_0^{\infty} (e^{\mathsf{T}}(z+tu))^2 g(t)t^{n-1}\,dt = |z|^2M_g(n-1)+2|z|(e^{\mathsf{T}}u)M_g(n)+(e^{\mathsf{T}}u)^2M_g(n+1).$$

By Lemma 3.2(b), we have here

$$(n+1)^2M_g(n)^2 \le n(n+2)M_g(n-1)M_g(n+1),$$

and so

$$|z|^2\frac{n(n+2)}{(n+1)^2}M_g(n-1) + 2|z|(e^{\mathsf{T}}u)M_g(n) + (e^{\mathsf{T}}u)^2M_g(n+1) \ge 0$$

(since the discriminant of this quadratic form is nonpositive). So

$$|z|^2M_g(n-1) + 2|z|(e^{\mathsf{T}}u)M_g(n) + (e^{\mathsf{T}}u)^2M_g(n+1) \ge \frac{1}{(n+1)^2}|z|^2M_g(n-1).$$

Substituting this in the integral, we get

$$1 \ge \frac{1}{(n+1)^2}|z|^2,$$

which proves the lemma. □

Lemma 3.14 *Let $f:\ \mathbb{R}^n \to \mathbb{R}$ be a logconcave function, let $F = \max f$, and let $L(t) = \{x \in \mathbb{R}^n :\ f(x) \ge t\}$. Then for $0 < s < t < F$,*

$$\frac{\mathrm{vol}(L(s))}{\mathrm{vol}(L(t))} \le \left(\frac{\ln(F/s)}{\ln(F/t)}\right)^n$$

Proof. Fix a point z where $f(z) = F$. Consider any point a on the boundary of $L(s)$. Let b be the intersection of the line through a and z with the boundary of $L(t)$. Let $b = ua + (1-u)z$. Since f is logconcave,

$$t \geq F^{1-u} s^u,$$

and so

$$u \geq \frac{\ln(F/t)}{\ln(F/s)}.$$

This means that $(\ln(F/s)/\ln(F/t))L(t)$ contains $L(s)$, and hence

$$\frac{\mathrm{vol}(L(s))}{\mathrm{vol}(L(t))} \leq \left(\frac{\ln(F/s)}{\ln(F/t)}\right)^n,$$

as claimed. □

Lemma 3.15 *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a logconcave function, let $F = \max f$, and let f_t denote the restriction of f to the level set $L(t) = \{x \in \mathbb{R}^n : f(x) \geq t\}$. Let $0 < s < t \leq F$ such that $t^{n+1} \leq s^n F$, and assume that f_t is isotropic. Then f_s is near-isotropic up to a factor of 6.*

Proof. Lemma 3.14 implies that

$$\int_{\mathbb{R}^n} f_s(x)\,dx \leq 3 \int_{\mathbb{R}^n} f_t(x)\,dx.$$

Hence for every unit vector u,

$$\begin{aligned}
\int_{\mathbb{R}^n} (u^T x)^2 \, d\pi_{f_s}(x) &= \frac{\int_{L(s)} (u^T x)^2 f(x)\,dx}{\int_{L(s)} f(x)\,dx} \\
&\geq \frac{\int_{L(t)} (u^T x)^2 f(x)\,dx}{\int_{L(s)} f(x)\,dx} \geq \frac{1}{3} \frac{\int_{L(t)} (u^T x)^2 f(x)\,dx}{\int_{L(t)} f(x)\,dx} \\
&= \frac{1}{3} \int_{\mathbb{R}^n} (u^T x)^2 \, d\pi_{f_t}(x).
\end{aligned}$$

On the other hand, Let L' be obtained by blowing up $L(t)$ from center z by a factor of $1 + 1/n$. Then $L(s) \subseteq L'$ by logconcavity, so

$$\begin{aligned}
&\int_{L(s)} (u^T x)^2 f(x)\,dx \leq \int_{L'} (u^T x)^2 f(x)\,dx \\
&\quad = \left(1 + \frac{1}{n}\right)^n \int_{L(t)} \left[u^T \left(\left(1 + \frac{1}{n}\right)x - \frac{1}{n} z\right)\right]^2 f\left(\left(1 + \frac{1}{n}\right)x - \frac{1}{n} z\right) dx.
\end{aligned}$$

Using that f decreases along semilines starting from z, we get

$$\begin{aligned}
&\int_{L(s)} (u^T x)^2 f(x)\,dx \\
&\quad \leq \left(1 + \frac{1}{n}\right)^n \int_{L(t)} \left[u^T \left(\left(1 + \frac{1}{n}\right)x - \frac{1}{n} z\right)\right]^2 f(x)\,dx
\end{aligned}$$

We can expand this into three terms:

$$\left(1+\frac{1}{n}\right)^{n+2}\int_{L(t)}(u^Tx)^2f(x)\,dx$$
$$-2\frac{1}{n}\left(1+\frac{1}{n}\right)^{n+1}\int_{L(t)}(u^Tx)(u^{\mathsf{T}}z)f(x)\,dx$$
$$+\frac{1}{n^2}\left(1+\frac{1}{n}\right)^{n}\int_{L(t)}(u^{\mathsf{T}}z)^2f(x)\,dx.$$

Here the middle term is 0 since f_t is isotropic, and the last term is

$$\begin{aligned}&\frac{1}{n^2}\left(1+\frac{1}{n}\right)^{n}\int_{L(t)}(u^{\mathsf{T}}z)^2f(x)\,dx\\ &\quad=\frac{1}{n^2}\left(1+\frac{1}{n}\right)^{n}(u^{\mathsf{T}}z)^2\int_{L(t)}f(x)\,dx\\ &\quad<\frac{e}{n^2}|z|^2\int_{L(t)}f(x)\,dx<3\int_{L(t)}f(x)\,dx\end{aligned}$$

by Lemma 3.13. The first term is

$$\left(1+\frac{1}{n}\right)^{n+2}\int_{L(t)}(u^Tx)^2f(x))\,dx<3\int_{L(t)}(u^Tx)^2f(x)\,dx,$$

and hence

$$\begin{aligned}\int_{\mathbb{R}^n}(u^Tx)^2\,d\pi_{f_s}(x) &= \frac{\int_{L(s)}(u^Tx)^2f(x)\,dx}{\int_{L(s)}f(x)\,dx}\le 3+3\frac{\int_{L(t)}(u^Tx)^2f(x)\,dx}{\int_{L(t)}f(x)\,dx}\\ &= 3+3\int_{\mathbb{R}^n}(u^Tx)^2\,d\pi_{f_t}(x)=6.\end{aligned}$$

□

We end this section with an important folklore theorem, generalizing Khinchine's inequality to logconcave functions.

Lemma 3.16 *For every* $\alpha\le\beta$,

$$\frac{\int_\alpha^\beta e^{-t}t^k\,dt}{\int_\alpha^\beta e^{-t}\,dt}\le k^k\cdot\left(\frac{\int_\alpha^\beta e^{-t}|t|\,dt}{\int_\alpha^\beta e^{-t}\,dt}\right)^k$$

Proof. Routine. □

Theorem 3.17 *If* X *is a random point from a logconcave distribution in* $\mathbb{R}^n$, *then*

$$\mathsf{E}(|X|^k)^{1/k}\le 2k\mathsf{E}(|X|).$$

Note that the Hölder inequality gives an opposite relation:

$$\mathsf{E}(|X|^k)^{1/k} \geq \mathsf{E}(|X|).$$

Proof. We can write this inequality as

$$\frac{\int_{\mathbb{R}^n} f(x)|x|^k\,dx}{\int_{\mathbb{R}^n} f(x)\,dx} \leq (2k)^k \cdot \left(\frac{\int_{\mathbb{R}^n} f(x)|x|\,dx}{\int_{\mathbb{R}^n} f(x)\,dx}\right)^k.$$

By Lemma 2.6 in [11], it suffices to prove that for any two points $a, b \in \mathbb{R}^n$ and $c \in \mathbb{R}$, we have

$$\frac{\int_0^1 e^{ct}|ta + (1-t)b|^k\,dt}{\int_0^1 e^{ct}\,dt} \leq (2k)^k \cdot \left(\frac{\int_0^1 e^{ct}|ta+(1-t)b|\,dx}{\int_0^1 e^{ct}\,dx}\right)^k.$$

Let c be the closest point of the segment $[a, b]$ to the origin. We can write any point on the segment $[a, b]$ as $c + s(b - a)$, where $\alpha = -|c-a|/|b-a| \leq s \leq \beta = |b - c|/|b - a|$. In this case,

$$|c + s(b - a)| \geq \max\{|c|, |s| \cdot |b - a|\}, \tag{11}$$

and of course

$$|c + s(b - a)| \leq |c| + |s| \cdot |b - a|.$$

Hence

$$\int_\alpha^\beta e^{c(s-\alpha)}|c + s(b - a)|^k\,ds \leq 2^{k-1}\int_\alpha^\beta e^{c(s-\alpha)}|c|^k\,ds + 2^{k-1}\int_\alpha^\beta e^{c(s-\alpha)}|s|^k|b-a|^k\,ds\,.$$

Here the first term is easy to evaluate, while the second term can be estimated using Lemma 3.16. We get that

$$\frac{\int_\alpha^\beta e^{c(s-\alpha)}|c + s(b - a)|^k\,ds}{\int_\alpha^\beta e^{c(s-\alpha)}\,ds} \leq 2^{k-1}|c|^k + 2^{k-1}|b-a|^k k^k \left(\frac{\int_\alpha^\beta e^{c(s-\alpha)}|s|\,ds}{\int_\alpha^\beta e^{c(s-\alpha)}\,ds}\right)^k = 2^{k-1}\left(\frac{\int_\alpha^\beta e^{c(s-\alpha)}|c|\,ds}{\int_\alpha^\beta e^{c(s-\alpha)}\,ds}\right)^k$$

$$+2^{k-1}k^k\left(\frac{\int_\alpha^\beta e^{c(s-\alpha)}|s||b-a|\,dt}{\int_\alpha^\beta e^{c(s-\alpha)}\,ds}\right)^k$$

$$\le (2k)^k\left(\frac{\int_\alpha^\beta e^{c(s-\alpha)}|c+s(b-a)|\,ds}{\int_\alpha^\beta e^{c(s-\alpha)}\,ds}\right)^k,$$

where the last step uses (11). □

4 Taming the function

To simplify notation, we introduce the following parameters:

$$R = 20\sqrt{n}\log n, \quad C = \frac{\varepsilon}{100}, \quad r = \frac{C^2}{5000\sqrt{n}}, \quad V_0 = \mathrm{vol}(rB) = \pi_n r^n.$$

4.1 Smoothing out

Next we define a "smoothed-out" version of the given density function f and prove its basic properties. Define

$$\hat{f}(x) = \min_C \frac{1}{\mathrm{vol}(C)}\int_C f(x+u)\,du,$$

where C ranges over all convex subsets of the ball $x + rB$ with $\mathrm{vol}(C) = V_0/16$. The quotient

$$\delta(x) = \frac{\hat{f}(x)}{f(x)}$$

is a certain measure of the smoothness of the function f at x.

The somewhat complicated definition of the function $\hat{f}$ serves to assure its logconcavity (Lemma 4.2). We'll also show that this function is not much smaller than f on the average (Lemma 4.3). The value of $\delta(x)$ governs the local smoothness of the function f in more natural ways than its definition (in the sequel to this paper [15] we'll show further evidence of this relevant to the hit-and-run walk). We start with a simple observation that shows that we could (at the cost of a factor of 2) replace equality in the condition on C by inequality:

Lemma 4.1 *For every convex subset $D \subseteq rB$ with $\mathrm{vol}(D) \ge \mathrm{vol}(B)/16$, we have*

$$\frac{1}{\mathrm{vol}(D)}\int_D f(x+u)\,du \ge \frac{1}{2}\hat{f}(x).$$

Proof. We can slice up D into convex sets $D = D_1 \cup \ldots \cup D_m$ so that $V_0/16 \le \mathrm{vol}(D_i) \le V_0/8$. For at least one i, we have

$$\frac{1}{\mathrm{vol}(D_i)}\int_{D_i} f(x+u)\,du \le \frac{1}{\mathrm{vol}(D)}\int_D f(x+u)\,du.$$

Let C be any convex subset of D_i of volume $\mathrm{vol}(rD)/16$, then

$$\frac{1}{\mathrm{vol}(C)}\int_C f(x+u)\,du \le \frac{2}{\mathrm{vol}(D_i)}\int_{D_i} f(x+u)\,du.$$

Since by definition

$$\frac{1}{\mathrm{vol}(C)}\int_C f(x+u)\,du \ge \hat{f}(x),$$

this proves the lemma. □

Lemma 4.2 *The function $\hat{f}$ is logconcave.*

Proof. For a fixed convex set $C \subseteq rB$, the function

$$f_C(x) = \int_C f(x+u)\,du$$

is the convolution of the function f with the characteristic function of the convex set $-C$, and so it is logconcave by Theorem 3.1. □

Lemma 4.3 *We have*

$$\int_{\mathbb{R}^n} \hat{f}(x)\,dx \ge 1 - 64r^{1/2}n^{1/4}.$$

To prove Lemma 4.3, we need a lemma from [10] (in a paraphrased form).

Lemma 4.4 *Let K be a convex set containing a ball of radius t. Let X be a uniform random point in K and let Y be a uniform random point at distance s from X. Then*

$$\mathsf{P}(Y \notin K) \le \frac{s\sqrt{n}}{2t}.$$

Proof. [of Lemma 4.3]. Consider the set $\{(X,T) \in \mathbb{R}^n \times \mathbb{R}_+ :\ T < f(X)\}$, and select a pair (X,T) randomly and uniformly from this set. Choose a uniform random point Z on the sphere of radius r around X. We estimate the probability that $T < f(Z)$.

First, fix X and Z, and then choose T. Since T is uniform in the interval $[0, f(X)]$, the probability that $T > f(Z)$ is

$$\mathsf{P}_T(T > f(Z)) = \begin{cases} 0, & \text{if } f(Z) > f(X),\\ 1 - \frac{f(Z)}{f(X)}, & \text{if } f(Z) \le f(X), \end{cases}$$

which we can also write as

$$\mathsf{P}_T(T > f(Z)) = \left|1 - \frac{f(Z)}{f(X)}\right|_+ .$$

Taking the expectation also over the choice of Z,

$$\begin{aligned} P_{T,Z}(T > f(Z)) &= \frac{1}{V_0}\int_{X+rB} P_T(T > f(z))\,dz \\ &= \frac{1}{V_0}\int_{X+rB} \left|1 - \frac{f(z)}{f(X)}\right|_+ dz. \end{aligned}$$

Let C be the convex set attaining the minimum in the definition of $\hat{f}(X)$, then it follows that

$$\begin{aligned} P_{T,Z}(T > f(Z)) &\geq \frac{1}{V_0}\int_C \left(1 - \frac{f(z)}{f(X)}\right) dz \\ &= \frac{1}{16}\left(1 - \frac{\hat{f}(X)}{f(X)}\right). \end{aligned}$$

Finally, taking expectation in X (which is from the distribution π_f), we get

$$P_{T,Z,X}(T < f(Z)) \leq \frac{1}{16}\int_{\mathbb{R}^n}\left(1 - \frac{\hat{f}(x)}{f(x)}f(x)\right) dx \tag{12}$$

$$= \frac{1}{16}\left(1 - \int_{\mathbb{R}^n}\hat{f}(x)\,dx\right). \tag{13}$$

Next, start by choosing T from its appropriate marginal distribution, and then choose X uniformly from $L(T)$, and then choose Z. Let F_0 be the smallest t for which the level set $L(t) = \{x \in \mathbb{R}^n : \ f(x) > t\}$ contains a ball of radius c_1 (which we'll choose later). If $T \leq F_0$, then $L(T)$ contains a ball of radius c_1, and so by Lemma 4.4, the probability that $Z \notin L(T)$ is at most $r\sqrt{n}/c_1$. Hence

$$\mathsf{P}(T > f(Z)) \leq \mathsf{P}(T > F_0) + \frac{r\sqrt{n}}{c_1}\mathsf{P}(T \leq F_0) = \mathsf{P}(T > F_0)(1 - \frac{r\sqrt{n}}{c_1}) + \frac{r\sqrt{n}}{c_1}.$$

Here $\mathsf{P}(T > F_0) \leq \pi_f(L(F_0)) \leq ec_1$ by Lemma 3.8, and so

$$\mathsf{P}(T > f(Z)) \leq ec_1 - er\sqrt{n} + \frac{r\sqrt{n}}{c_1}.$$

The best choice of c_1 is $c_1 = r^{1/2}n^{1/4}e^{-1/2}$, which gives

$$\mathsf{P}(T > f(Z)) \leq 2e^{1/2}r^{1/2}n^{1/4}.$$

Together with (12), this proves the lemma. □

This lemma implies that the distributions π_f and $\pi_{\hat{f}}$ are close:

Corollary 4.5 *The total variation distance of π_f and $\pi_{\hat{f}}$ is less than $100r^{1/2}n^{1/4}$.*

Proof. Let $c = 64r^{1/2}n^{1/4}$. The density function of $\pi_{\hat{f}}$ is $\hat{f}/(1-c)$, and hence (using that $\hat{f} \le f$ and $c < 1/5$),

$$\begin{aligned} d_{\mathrm{tv}}(\pi_f, \pi_{\hat{f}}) &= \int_{\mathbb{R}^n} \left| \frac{\hat{f}(x)}{1-c} - f(x) \right|_+ dx \le \int_{\mathbb{R}^n} \frac{f(x)}{1-c} - f(x)\, dx = \frac{c}{1-c} \\ &< 80r^{1/2}n^{1/4}. \end{aligned}$$

□

Using the function $\hat{f}$, we can state and prove a bound on the smoothness of the original function, at least in an average sense.

Lemma 4.6 *Let $x \in \mathbb{R}^n$, and let U be a uniform random point in $x + rB$. Then with probability at least* $15/16$,

$$f(U) \le 2\frac{f(x)^2}{\hat{f}(x)}.$$

Proof. Consider the set

$$S = \left\{ u \in x + rB : \; f(u) > 2\frac{f(x)^2}{\hat{f}(x)} \right\}.$$

Clearly S is convex, and so is the set S' obtained by reflecting S in x. Furthermore, for every $y \in S'$ we have by logconcavity

$$f(y)f(2x - y) \le f(x)^2,$$

and since $f(2x - y) > 2f(x)^2/\hat{f}(x)$ by definition, we have $f(y) < \frac{1}{2}\hat{f}(x)$. By Lemma 4.1, this can only happen on a convex set of measure less than $1/16$, which proves the lemma. □

5 Geometric distance and probabilistic distance.

For any point u, let P_u be the distribution obtained on taking one ball step from u. Here we show that if two points are close in a geometric sense, then the distributions obtained after making one step of the random walk from them are also close in total variation distance.

Lemma 5.1 *Let $u, v \in \mathbb{R}^n$ such that*

$$d(u, v) < \frac{r}{8\sqrt{n}}.$$

Then for the ball walk,

$$d_{tv}(P_u, P_v) < 1 - \frac{1}{16}\delta(u).$$

Proof. Let B_u, B_v be the balls of radius r around u and v respectively. Let $C = B_u \cap B_v$ and C' be the subset of C where $f(x) \leq \bar{f} = \min f(u), f(v)$. Then for any point $x \in C'$, the probability density of going from u to x or v to x is at least

$$\frac{1}{\text{vol}(B(r))} \frac{f(x)}{\bar{f}}$$

Thus,

$$d_{tv}(P_u, P_v) < 1 - \frac{1}{\text{vol}(B(r))} \int_{x \in C'} \frac{f(x)}{\bar{f})} dx.$$

Since $d(u,v) \leq r/8\sqrt{n}$,

$$\text{vol}(C') \geq \frac{1}{4}\text{vol}(C) \geq \frac{1}{8}\text{vol}(B(r)) \geq c_0 \text{vol}(B(r)).$$

By Lemma 4.1,

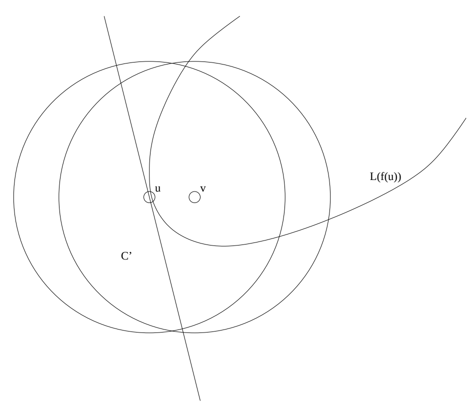

Figure 1: The proof of Lemma 5.1.

$$\int_{x \in C'} f(x)dx \geq \frac{1}{2}\hat{f}(u)\text{vol}(C') \geq \frac{1}{16}\hat{f}\ (u)\text{vol}(B(r)).$$

Hence,

$$d_{tv}(P_u, P_v) < 1 - \frac{1}{16}\frac{\hat{f}(u)}{\bar{f}} = 1 - \frac{1}{16}\frac{\hat{f}(u)}{f(u)} = 1 - \frac{1}{16}\delta(u).$$

□

6 Proof of the mixing bound.

Let $K = S_1 \cup S_2$ be a partition into measurable sets with $\pi_f(S_1), \pi_f(S_2) > \varepsilon$. We will prove that

$$\int_{S_1} P_u(S_2)\, d\pi_f \quad \geq \quad \frac{r}{2^{10}n}(\pi_f(S_1) - \varepsilon)(\pi_f(S_2) - \varepsilon) \tag{14}$$

We can read the left hand side as follows: we select a random point X from distribution π and make one step to get Y. What is the probability that $X \in S_1$ and $Y \in S_2$? It is well known that this quantity remains the same if S_1 and S_2 are interchanged.

For $i \in \{1, 2\}$, let

$$\begin{aligned} S_i' &= \{x \in S_i : \ P_x(S_{3-i}) < \frac{1}{32}\delta(x), \\ S_3' &= K \setminus S_1' \setminus S_2'. \end{aligned}$$

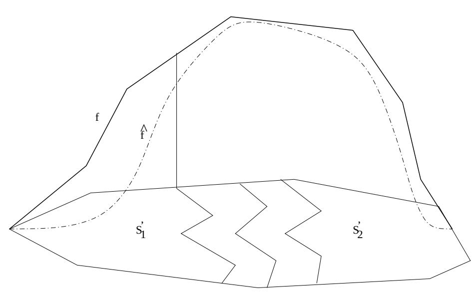

Figure 2: The mixing proof.

First, suppose that $\pi_{\hat f}(S_1') \le \pi_{\hat f}(S_1)/2$. Then the left hand side of (14) is at least

$$\frac{1}{32}\int_{u\in S_1\setminus S_1'} \frac{\hat f(u)}{f(u)} f(u)\,du = \frac{1}{32}\pi_{\hat f}(S_1\setminus S_1') \ge \frac{1}{64}\pi_{\hat f}(S_1).$$

Lemma 4.3 implies that

$$\pi_{\hat f}(S_1) \ge \pi_f(S_1) - \varepsilon.$$

Hence,

$$\int_{S_1} P_u(S_2)\,d\pi_f \ge \frac{1}{64}(\pi_f(S_1)-\varepsilon)$$

which implies (14).

So we can assume that $\pi_f(S_1') \ge \pi_f(S_1)/2$, and similarly $\pi_f(S_2') \ge \pi_f(S_2)/2$. By Lemma 5.1, for any two points $u_1 \in S_1'$, $u_2 \in S_2'$,

$$d(u,v) \quad \ge \quad \frac{r}{8\sqrt{n}} \tag{15}$$

Applying Theorem 2.3 to $\hat f$, we get

$$\pi_{\hat f}(S_3') \ge \frac{r}{16n}\pi_{\hat f}(S_1')\pi_{\hat f}(S_2') \ge \frac{r}{16n}(\pi_f(S_1)-\varepsilon)(\pi_f(S_2)-\varepsilon).$$

Therefore,

$$\begin{aligned}\int_{S_1} P_u(S_2)\,d\pi_f \quad &\ge \quad \frac{1}{2}\int_{S_3'} \frac{\hat f(u)}{32 f(u)} f(u)du \\ &\ge \quad \frac{1}{64}\pi_{\hat f}(S_3') \\ &\ge \quad \frac{r}{2^{10}n}(\pi_f(S_1)-\varepsilon)(\pi_f(S_2)-\varepsilon)\end{aligned}$$

and (14) is proved.

By Corollary 1.5 in [14] it follows that for all $m \ge 0$, and every measurable set S,

$$|\sigma^m(S) - \pi_f(S)| \le 2\varepsilon + \exp\left(-\frac{mr^2}{2^{42}n^2}\right),$$

which proves Theorem 2.2.

Acknowledgement. We are indebted to Keith Ball for his kind advice on the geometric aspects of logconcave functions.

References

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

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

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

[4] M. Dyer, A. Frieze and R. Kannan: A random polynomial time algorithm for estimating volumes of convex bodies, *Proc. 21st Annual ACM Symp. on the Theory of Computing* (1989), 375–381.

[5] M. Dyer, A. Frieze and L. Stougie: talk at the Math. Prog. Symp., Ann Arbor, Mich. (1994).

[6] 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.

[7] A. Frieze and R. Kannan: Log-Sobolev inequalities and sampling from log-concave distributions, To appear in the Annals of Applied Probability

[8] A. Kalai and S. Vempala: Efficient algorithms for universal portfolios, in: *Proc. 41st Annual IEEE Symp. on the Foundations of Computer Science* (2000), 486-491.

[9] R. Kannan and L. Lovász: A logarithmic Cheeger inequality and mixing in random walks, in: *Proc. 31st Annual ACM STOC* (1999), 282-287.

[10] 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.

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

[12] L. Lovász (1992): How to compute the volume? *Jber. d. Dt. Math.-Verein, Jubiläumstagung 1990*, B. G. Teubner, Stuttgart, 138–151.

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

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

[15] L. Lovász and S. Vempala: Hit-and-run is fast and fun.

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

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

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

[19] A. Prékopa: *Stochastic Programming*, Akadémiai Kiadó, Budapest and Kluwer, Dordrecht 1995.

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

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

[22] Z.B. Zabinsky, R.L. Smith, J.F. McDonald, H.E. Romeijn and D.E. Kaufman: Improving hit-and-run for global optimization, *J. Global Optim.* **3** (1993), 171–192.