Stochastic Localization + Stieltjes Barrier = Tight Bound for Log-Sobolev

Yin Tat Lee[*], Santosh S. Vempala[†]

November 15, 2017

Abstract

Logarithmic Sobolev inequalities are a powerful way to estimate the rate of convergence of Markov chains and to derive concentration inequalities on distributions. We prove that the log-Sobolev constant of any isotropic logconcave density in $\mathbb{R}^n$ with support of diameter D is $\Omega(1/D)$, resolving a question posed by Frieze and Kannan in 1997. This is asymptotically the best possible estimate and improves on the previous bound of $\Omega(1/D^2)$ by Kannan-Lovász-Montenegro. It follows that for any isotropic logconcave density, the ball walk with step size $\delta = \Theta(1/\sqrt{n})$ mixes in $O\left(n^2 D\right)$ proper steps from *any* starting point. This improves on the previous best bound of $O(n^2D^2)$ and is also asymptotically tight. The new bound leads to the following refined large deviation inequality for a L-Lipschitz function g over an isotropic logconcave density p: for any $t > 0$,

$$\mathbb{P}_{x\sim p}\left(|g(x) - \bar{g}| \geq c \cdot L \cdot t\right) \leq \exp\left(-\frac{t^2}{t + \sqrt{n}}\right)$$

where $\bar{g}$ is the median or mean of g for $x \sim p$; this generalizes/improves on previous bounds by Paouris and by Guedon-Milman. Our main proof is based on stochastic localization together with a Stieltjes-type barrier function.

1 Introduction

This purpose of this paper is to understand the asymptotic behavior of the log-Sobolev and log-Cheeger constants of convex bodies and logconcave distributions in $\mathbb{R}^n$. These fundamental parameters, which we will define presently, have many important connections and applications (cf. [15]). To introduce them, we first remind the reader of the Cheeger constant (a.k.a. isoperimetric constant or expansion).

Definition 1. For a density p in $\mathbb{R}^n$, the Cheeger constant of p is defined as

$$\psi_p \stackrel{\text{def}}{=} \inf_{S\subseteq\mathbb{R}^n} \frac{\int_{\partial S} p(x)dx}{\min\left\{\int_S p(s)dx, \int_{\mathbb{R}^n\setminus S} p(x)dx\right\}}.$$

Kannan, Lovász and Simonovits [11] conjectured that for any logconcave density, the Cheeger constant satisfies[1] $\psi_p \gtrsim \|A\|_{\text{op}}^{-1/2}$ where A is the covariance matrix of p. A density/distribution is called *isotropic* if its covariance matrix is the identity, a normalization that can be achieved via an affine transformation. For isotropic logconcave densities, the conjecture says the Cheeger constant is $\Omega(1)$. The current best estimate of ψ_p is the following recent result.

Theorem 2 ([18]). *For any logconcave density p in $\mathbb{R}^n$ with covariance matrix A,*

$$\psi_p \gtrsim \left(\text{Tr}\left(A^2\right)\right)^{-1/4}.$$

[*]Microsoft Research and University of Washington, yile@microsoft.com and yintat@uw.edu

[†]Georgia Tech, vempala@gatech.edu

[1]We write $\gtrsim$to denote "at least a constant times".

For isotropic p, this gives a bound of $\psi_p \gtrsim n^{-\frac{1}{4}}$. The KLS hyperplane conjecture plays a central role in asymptotic convex geometry, and implies several other well-known (older) conjectures, including slicing (or hyperplane) and thin-shell (or variance) conjectures, the Poincare conjecture, central limit, exponential concentration etc. (see e.g., [4]).

The KLS conjecture was motivated by the study of the convergence of a Markov chain, the *ball walk* in a convex body. To sample uniformly from a convex body, the ball walk starts at some point in the body, picks a random point in the ball of radius δ around the current point and if the chosen point is in the body, it steps to the new point. It can be generalized to sampling any logconcave density by using a Metropolis filter. As shown in [12], the ball walk applied to a logconcave density mixes in $O^*(n^2/\psi_p^2)$ steps from a warm start, which using the current-best bound [18] is $O^*(n^{2.5})$. Looking closer, from a starting distribution Q_o, the distance of the distribution obtained after t steps from Q_o to the stationary distribution Q drops as

$$d(Q_t, Q) \le d(Q_0, Q)\left(1 - \frac{\phi^2}{2}\right)^t$$

where ϕ is the conductance of the Markov chain and $d(.,.)$ is the χ-squared distance. The conductance can be viewed as the Cheeger constant of the Markov chain. Thus the number of steps needed is $O(\phi^{-2}\log(1/d(Q_0, Q)))$. Roughly speaking, for the ball walk applied to a logconcave density p, the conductance is $\Omega(\psi_p/n)$, leading to the bound of $O^*(n^2/\psi_p^2)$ steps from a warm start. The dependence on the starting distribution leads to an additional factor of n in the running time when the starting distribution is not warm (i.e., $d(Q_1, Q)$ after one step can be $e^{\tilde{\Omega}(n)}$). This is a general issue for Markov chains. One way to address this is via the log-Sobolev constant [6, 15]. We first define it for a density, then for a Markov chain.

Definition 3. For a density p the log-Sobolev constant ρ_p is the largest ρ such that for every smooth function $f : \mathbb{R}^n \to \mathbb{R}$ with $\int f^2 dp = 1$, we have

$$\rho \le \frac{2\int \|\nabla f\|^2 dp}{\int f^2 \log f^2 dp}.$$

A closely related parameter is the following.

Definition 4. The *log-Cheeger* or log-isoperimetric constant κ_p of a density p in $\mathbb{R}^n$ is

$$\kappa_p = \inf_{S \subseteq \mathbb{R}^n} \frac{p(\partial S)}{\min\{p(S), p(\mathbb{R}^n \setminus S)\}\sqrt{\log\left(\frac{1}{p(S)}\right)}}.$$

It is known that $\rho_p = \Theta(\kappa_p^2)$ (see e.g., [14]). The log-Cheeger constant shows more explicitly that the log-Sobolev constant is a uniform bound on the expansion "at every scale".

For a reversible Markov chain with transition operator P and stationary density Q, the analogous definition is the infimum over all smooth functions satisfying $f : \mathbb{R}^n \to \mathbb{R}$ with $\int f^2 dp = 1$ of

$$\rho(P) = \frac{\int_x \int_y \|f(x) - f(y)\|^2 P(x,y) dQ(x)}{\int f^2 \log f^2 dQ}.$$

Diaconis and Saloff-Coste [6] show that the distribution after t steps satisfies $\mathrm{Ent}(Q_t) \le e^{-4\rho(P)t}\mathrm{Ent}(Q_0)$ where $\mathrm{Ent}(Q_t) = \int Q_t \log \frac{Q_t}{Q_0} dQ_0$ is the entropy with respect to the stationary distribution. Thus, the dependence of the mixing time on the starting distribution goes down from $\log(1/d(Q_0, Q))$ to $\log\log(1/d(Q_0, Q)))$. Moreover, just as in the case of the Cheeger constant, for the ball walk, the Markov chain parameter is determined by the log-Sobelov constant ρ_p for sampling from the density p. It is thus natural to ask for the best possible bound on ρ_p or κ_p. Unlike the Cheeger constant, which is conjectured to be at least a constant for isotropic logconcave densities, it is known that ρ_p cannot be bounded from below by a universal constant, in particular for distributions that are not "ψ_2" (distributions with sub-Gaussian tail).

Kannan, Lovász and Montenegro [13] gave the following bound on κ_p. Our main result (Theorem 8) is an improvement of this bound to the best possible.

Theorem 5 ([13]). *For an isotropic logconcave density $K \subset \mathbb{R}^n$ with support of diameter D, we have $\kappa_p \gtrsim \frac{1}{D}$ and $\rho_p \gtrsim \frac{1}{D^2}$.*

From the above bound, it follows that the ball walk mixes in $O\left(n^2 D^2\right)$ *proper* steps of step size δ. A proper step is one where the current point changes. For an isotropic logconcave density $\delta = \Theta(\frac{1}{\sqrt{n}})$ is small enough so that the number of wasted steps is of the same order as the number of proper steps in expectation. Moreover, by restricting to a ball of radius $D = O(\sqrt{n})$, the resulting distribution remains near-isotropic and very close in total variation distance to the original. Together, this considerations imply a bound of $O^*(n^3)$ proper steps from any starting point as shown in [13]. Is this bound the best possible? From a warm start, the KLS conjecture implies a bound of $O^*(n^2)$ steps and current best bound is $O^*(n^{2.5})$. Thus, the mixing of the ball walk, which was the primary motivation for formulation of the KLS conjecture, also provides a compelling reason to study the log-Sobolev constant. Estimating the log-Sobolev constant was posed as an open problem by Frieze and Kannan [8] when they analyzed the log-Sobolev constant of the grid walk to sample sufficiently smooth logconcave densities.

The Cheeger and log-Sobolev constants also play an important role in the phenomenon known as concentration of measure. The following result is due to Gromov and Milman.

Theorem 6 (Lipschitz concentration [9]). *For any L-Lipschitz function g in $\mathbb{R}^n$, and isotropic logconcave density p,*

$$\mathbb{P}_{x\sim p}\left(|g(x) - \mathbb{E}g| \geq c \cdot L \cdot t\right) \leq e^{-t\psi_p}.$$

Using the best-known estimate of the Cheeger constant gives a bound of $e^{-t/n^{1/4}}$ [18]. A different bound, independent of the Cheeger constant, for the deviation in length of a random vector was given in a celebrated paper by Paouris [20] and improved by Guedon and Milman [10] (Paouris' result has only the second term in the minimum below, and is sharp when $t \gtrsim \sqrt{n}$).

Theorem 7 ([10, 20]). *For any isotropic logconcave density p,*

$$\mathbb{P}_{x\sim p}\left(\left|\|x\| - \sqrt{n}\right| \geq c \cdot t\right) \leq e^{-\min\left\{\frac{t^3}{n}, t\right\}}.$$

Our tight log-Sobolev bound will be useful in proving an improved concentration inequality.

1.1 Results

Our main theorem is the following.

Theorem 8. *For any isotropic logconcave density p with support of diameter D, the log-Cheeger constant satisfies $\kappa_p \gtrsim \frac{1}{\sqrt{D}}$ and the log-Sobolev constant satisfies $\rho_p \gtrsim \frac{1}{D}$.*

As we show in Section 6, these bounds are the best possible (Lemma 32). The improved bound has interesting consequences. The first is an improved concentration of mass inequality for logconcave densities. In particular, this gives an alternative proof of Paouris' (optimal) inequality [20] for the large deviation case ($t \geq \sqrt{n}$).

Theorem 9. *For any L-Lipschitz function g in $\mathbb{R}^n$ and any isotropic logconcave density p, we have that*

$$\mathbb{P}_{x\sim p}\left(|g(x) - \bar{g}(x)| \geq c \cdot L \cdot t\right) \leq \exp\left(-\frac{t^2}{t + \sqrt{n}}\right)$$

where $\bar{g}$ is the median or mean of $g(x)$ for $x \sim p$.

As mentioned earlier, the previous best bound is $\exp(-\min(\frac{t^3}{n}, t))$ [10] for the function $g(x) = \|x\|$ and $\exp(-\frac{t}{n^{1/4}})$ for a general Lipschitz function g [18]. The new bound can be viewed as an improvement and generalization of both (note that we can write the exponent as $-\min(\frac{t^2}{n}, t)$). Also this concentration result does not need bounded support for the density p.

As a second consequence, we circle back to the analysis of the ball walk to resolve the open problem posed by Frieze and Kannan [8].

Theorem 10. *The ball walk with step size $\delta = \Theta(1/\sqrt{n})$ applied to an isotropic logconcave density in $\mathbb{R}^n$ with support of diameter D mixes in $O^*(n^2 D)$ proper steps from any starting point of positive density (or from any interior point of a convex body).*

The choice of $\delta = \Theta\left(1/\sqrt{n}\right)$ is the best possible for isotropic logconcave distributions (Lemma 33). The bound on the number of steps improves on the previous best bound of $O^*(n^2 D^2)$ proper steps for the mixing of the ball walk from an arbitrary starting point [13] and as we show in Section 6, $O(n^2 D)$ is the best possible bound. For sampling, we can restrict the density to a ball of radius $O(\sqrt{n})$ losing only a negligibly small measure, so the bound is $O(n^{2.5})$ from an arbitrary starting point, which matches the current best bound from a warm start. The mixing time from a warm start for an isotropic logconcave density is $O(n^2 \psi_p^2)$, or $O(n^2)$ if the KLS conjecture is true; but from an arbitrary start, our analysis is essentially the best possible, independent of any further progress on the conjecture!

2 Approach: Stochastic localization

In this section, we describe the stochastic localization method introduced by Eldan [7], and, in particular, the variant of the method used in [18]. The idea of the approach is to gradually modify the density p by making infinitesimal changes so that the measure of a set and of its boundary is not changed by much, but the density function itself accumulates a significant Gaussian component, i.e., it looks like a Gaussian density function times a logconcave function. For such densities, we can use standard localization (or other methods) to prove that the log-Sobolev constant is large. While this is the same high-level approach as in previous papers, several new challenges arise. First, unlike previous applications, we cannot simply work with subsets of measure 1/2 or a constant, it is crucial to consider arbitrarily small subsets. Second, as we will discuss presently, we need a more refined potential function to understand the evolution of the measure.

Definition 11. Given a logconcave distribution p, we define the following stochastic differential equation:

$$c_0 = 0, \quad dc_t = dW_t + \mu_t dt, \tag{2.1}$$

where the probability distribution p_t, the mean μ_t and the covariance A_t are defined by

$$p_t(x) = \frac{e^{c_t^T x - \frac{t}{2}\|x\|_2^2} p(x)}{\int_{\mathbb{R}^n} e^{c_t^T y - \frac{t}{2}\|y\|_2^2} p(y) dy}, \quad \mu_t = \mathbb{E}_{x \sim p_t} x, \quad A_t = \mathbb{E}_{x \sim p_t} (x - \mu_t)(x - \mu_t)^T.$$

We collect the properties of this stochastic localization that we will use in this paper in the following Lemma.

Lemma 12 ([18])**.** *For any logconcave distribution p with bounded support, the stochastic process p_t defined in Definition 11 exists and is unique. Also, p_t is a martingale, and in particular, for any $x \in \mathbb{R}^n$*

$$dp_t(x) = (x - \mu_t)^T dW_t \cdot p_t(x).$$

Its covariance matrix satisfies

$$dA_t = \int_{\mathbb{R}^n} (x - \mu_t)(x - \mu_t)^T ((x - \mu_t)^T dW_t) \cdot p_t(x) dx - A_t^2 dt. \tag{2.2}$$

In previous papers [7, 18], the spectral norm of the covariance the $\|A_t\|_{\text{op}}$ is bounded via a potential function of the form $\operatorname{Tr}\left(A_t^q\right)$. However, to obtain a tight result without extraneous logarithmic factors, we study the Stieltjes-type potential $\operatorname{Tr}\left((uI - A_t)^{-q}\right)$.

To define the potential, fix integers $n, q \geq 1$ and a positive number $\Phi > 0$. Let $u(X)$ be the real-valued function on $n \times n$ symmetric matrices defined by the solution of the following equation

$$\operatorname{Tr}((uI - X)^{-q}) = \Phi \text{ and } X \preceq uI \tag{2.3}$$

Note that this is the same as the solution to $\sum_{i=1}^n \frac{1}{(u-\lambda_i)^q} = \Phi$ and $\lambda_i \leq u$ for all i where λ_i are the eigenvalues of X. Similar potentials have been used to analyze empirical covariance estimation [21], to build graph sparsifiers [3, 1, 16, 17] and to solve bandit problems [2].

The proof has the following ingredients:

1. We show that for time t up to $O(n^{-\frac{1}{2}})$, the spectral norm of the covariance stays bounded (by a constant, say 2) with large probability (Lemma 19). This requires the use of the Stieltjes-type potential function.
2. Then we consider any measurable subset S, with $g_0 = p_0(S)$ and analyze its measure at time t, i.e., $g_t = p_t(S)$. In particular we show that up to time $(\log g_0 + D)^{-1}$, the expectation of $g_t\sqrt{\log(1/g_t)}$ remains large, i.e., a constant factor times its initial value (Lemma 28).
3. The density at time t has a Gaussian component of variance $1/t$. For such a distribution, the log-Cheeger constant is $\Omega(\sqrt{t})$ (Theorem 29).

Together these facts will imply the main theorem.

3 Preliminaries

In this section, we review some basic definitions and theorems that we use in the proofs.

3.1 Stochastic calculus

Given real-valued stochastic processes x_t and y_t, the quadratic variations $[x]_t$ and $[x,y]_t$ are real-valued stochastic processes defined by

$$[x]_t = \lim_{|P|\to 0}\sum_{n=1}^{\infty}\left(x_{\tau_n} - x_{\tau_{n-1}}\right)^2 \quad \text{and} \quad [x,y]_t = \lim_{|P|\to 0}\sum_{n=1}^{\infty}\left(x_{\tau_n} - x_{\tau_{n-1}}\right)\left(y_{\tau_n} - y_{\tau_{n-1}}\right),$$

where $P = \{0 = \tau_0 \le \tau_1 \le \tau_2 \le \cdots \uparrow t\}$ is a stochastic partition of the non-negative real numbers, $|P| = \max_n (\tau_n - \tau_{n-1})$ is called the *mesh* of P and the limit is defined using convergence in probability. Note that $[x]_t$ is non-decreasing with t and $[x,y]_t$ can be defined as

$$[x,y]_t = \frac{1}{4}\left([x+y]_t - [x-y]_t\right).$$

If the processes x_t and y_t satisfy the SDEs $dx_t = \mu(x_t)dt + \sigma(x_t)dW_t$ and $dy_t = \nu(y_t)dt + \eta(y_t)dW_t$ where W_t is a Wiener process, we have that $[x]_t = \int_0^t \sigma^2(x_s)ds$ and $[x,y]_t = \int_0^t \sigma(x_s)\eta(y_s)ds$ and $d[x,y]_t = \sigma(x_t)\eta(y_t)dt$; for a vector-valued SDE $dx_t = \mu(x_t)dt + \Sigma(x_t)dW_t$ and $dy_t = \nu(y_t)dt + M(y_t)dW_t$, we have that $[x^i, x^j]_t = \int_0^t (\Sigma(x_s)\Sigma^T(x_s))_{ij}ds$ and $d[x^i, y^j]_t = (\Sigma(x_t)M^T(y_t))_{ij}dt$.

Lemma 13 (Itô's formula). *Let x be a semimartingale and f be a twice continuously differentiable function, then*

$$df(x_t) = \sum_i \frac{df(x_t)}{dx^i}dx^i + \frac{1}{2}\sum_{i,j}\frac{d^2 f(x_t)}{dx^i dx^j}d[x^i, x^j]_t.$$

The next two lemmas are well-known facts about Wiener processes; first the reflection principle.

Lemma 14 (Reflection principle). *Given a Wiener process $W(t)$ and $a, t \ge 0$, then we have that*

$$\mathbb{P}(\sup_{0\le s\le t} W(s) \ge a) = 2\mathbb{P}(W(t) \ge a).$$

Second, a decomposition lemma for continuous martingales.

Theorem 15 (Dambis, Dubins-Schwarz theorem). *Every continuous local martingale M_t is of the form*

$$M_t = M_0 + W_{[M]_t} \text{ for all } t \ge 0$$

where W_s is a Wiener process.

3.2 Logconcave functions and isoperimetry

We say a logconcave distribution is *nearly* isotropic if its covariance matrix A satisfies $\Omega(1)\cdot I \preceq A \preceq O(1)\cdot I$. The following lemma about logconcave densities is folklore, see e.g., [19].

Lemma 16 (Logconcave moments)**.** *Given a logconcave density p in $\mathbb{R}^n$, and any integer $k \geq 1$,*

$$\mathbb{E}_{x\sim p}\left\|x\right\|^{k} \leq (2k)^{k}\left(\mathbb{E}_{x\sim p}\left\|x\right\|^{2}\right)^{k/2}.$$

Theorem 17 (Poincaré constant [22, 5])**.** *For any logconcave density p in $\mathbb{R}^n$ and any function g in $\mathbb{R}^n$, we have*

$$Var_{p}\left(g(x)\right) \lesssim \psi_{p}^{-2}\cdot\mathbb{E}_{p}\left(\left\|\nabla g(x)\right\|_{2}^{2}\right)$$

4 Main proof

The goal of this section is to prove the following theorem, which implies Theorem 8.

Theorem 18. *Given an isotropic logconcave distribution p with support of diameter D. Then, for any measurable subset S,*

$$p(\partial S) \geq \Omega\left(\frac{\log\frac{1}{p(S)}}{D}+\sqrt{\frac{\log\frac{1}{p(S)}}{D}}\right)p(S).$$

4.1 Bounding the spectral norm of the covariance matrix

The main lemma of this section is the following.

Lemma 19. *There is some universal constant $c \geq 0$ such that for any $0 \leq T \leq \frac{1}{c\cdot\sqrt{n}}$, we have that*

$$\mathbb{P}(\max_{t\in[0,T]}\left\|A_{t}\right\|_{2}\geq 2)\leq 2\exp(-\frac{1}{cT}).$$

We defer the proof of the following lemma to the end of this section.

Lemma 20. *We have that*

$$du(A_{t})=\alpha_{t}^{T}dW_{t}+\beta_{t}dt$$

where

$$\begin{aligned}
\alpha_{t} =& \frac{1}{\kappa_{t}}\mathbb{E}_{x\sim\tilde{p}_{t}}x^{T}(uI-A_{t})^{-(q+1)}x\cdot x,\\
\frac{\beta_{t}}{q+1} \leq& \frac{1}{2\kappa_{t}}\mathbb{E}_{x,y\sim\tilde{p}_{t}}x^{T}(uI-A_{t})^{-1}y\cdot x^{T}(uI-A_{t})^{-(q+1)}y\cdot x^{T}y\\
&-\frac{1}{\kappa_{t}^{2}}\mathbb{E}_{x,y\sim\tilde{p}_{t}}x^{T}(uI-A_{t})^{-(q+1)}x\cdot y^{T}(uI-A_{t})^{-(q+2)}y\cdot x^{T}y\\
&+\frac{1}{2\kappa_{t}^{3}}\mathrm{Tr}((uI-A_{t})^{-(q+2)})\cdot\mathbb{E}_{x,y\sim\tilde{p}_{t}}x^{T}(uI-A_{t})^{-(q+1)}x\cdot y^{T}(uI-A_{t})^{-(q+1)}y\cdot x^{T}y,\\
\kappa_{t} =& \mathrm{Tr}((uI-A_{t})^{-(q+1)})
\end{aligned}$$

and $\tilde{p}_t$ be the translation of p_t defined by $\tilde{p}_t(x) = p_t(x+\mu_t)$.

To estimate α_t, we need the following lemma proved in [18].

Lemma 21 ([18, Lemma 25])**.** *Given a logconcave distribution p with mean μ and covariance A. For any $C \succeq 0$, we have that*

$$\left\|\mathbb{E}_{x\sim p}(x-\mu)^{T}C(x-\mu)(x-\mu)\right\|_{2}=O(\left\|A\right\|_{\mathrm{op}}^{1/2})\mathrm{Tr}\left(A^{1/2}CA^{1/2}\right).$$

To estimate β_t, we prove the following bound that crucially uses the KLS bound for non-isotropic log-concave distribution (Theorem 2).

Lemma 22. *Given a logconcave distribution p with mean μ and covariance A. For any $B^{(1)}, B^{(2)}, B^{(3)} \succeq 0$,*

$$\begin{aligned}&\left|\mathbb{E}_{x,y\sim p}(x-\mu)^T B^{(1)}(y-\mu)\cdot(x-\mu)^T B^{(2)}(y-\mu)\cdot(x-\mu)^T B^{(3)}(y-\mu)\right|\\ \le& O(1)\cdot \mathrm{Tr}(A^{\frac{1}{2}}B^{(1)}A^{\frac{1}{2}})\cdot\left\|A^{\frac{1}{2}}B^{(2)}A^{\frac{1}{2}}\right\|_F\cdot\left\|A^{\frac{1}{2}}B^{(3)}A^{\frac{1}{2}}\right\|_{\mathrm{op}}.\end{aligned}$$

Proof. Without loss of generality, we can assume p is isotropic. Furthermore, we can assume $B^{(1)}$ is diagonal. Let $\Delta^{(k)} = \mathbb{E}_{x\sim p} xx^T \cdot x^T e_k$. Then, we have that

$$\begin{aligned}\mathbb{E}_{x,y\sim p} x^T B^{(1)} y\cdot x^T B^{(2)} y \cdot x^T B^{(3)} y &= \sum_k B^{(1)}_{kk}\mathrm{Tr}(B^{(2)}\Delta^{(k)}B^{(3)}\Delta^{(k)})\\ &\le \left\|B^{(3)}\right\|_{\mathrm{op}}\sum_k B^{(1)}_{kk}\mathrm{Tr}(\Delta^{(k)}B^{(2)}\Delta^{(k)}).\end{aligned} \tag{4.1}$$

Note that

$$\begin{aligned}\mathrm{Tr}(\Delta^{(k)}B^{(2)}\Delta^{(k)}) &= \mathbb{E}_{x\sim p} x^T B^{(2)}\Delta^{(k)} x\cdot x_k\\ &\le \sqrt{\mathbb{E}x_k^2}\sqrt{\mathrm{Var}_{x\sim p} x^T B^{(2)}\Delta^{(k)} x}\\ &= \sqrt{\mathrm{Var}_{y\sim\tilde{p}} y^T \left(B^{(2)}\right)^{\frac{1}{2}}\Delta^{(k)}\left(B^{(2)}\right)^{-\frac{1}{2}} y}\end{aligned}$$

where $\tilde{p}$ is the distribution given by $\left(B^{(2)}\right)^{\frac{1}{2}} x$ where $x\sim p$. Note that $\tilde{p}$ is a logconcave distribution with mean 0 and covariance $B^{(2)}$. Theorem 17 together with Theorem 2 shows that

$$\mathrm{Var}_{y\sim\tilde{p}} y^T\left(B^{(2)}\right)^{\frac{1}{2}}\Delta^{(k)}\left(B^{(2)}\right)^{-\frac{1}{2}} y \le O\left(\left\|B^{(2)}\right\|_F\right)\cdot\mathbb{E}_{y\sim\tilde{p}}\left\|\left(B^{(2)}\right)^{\frac{1}{2}}\Delta^{(k)}\left(B^{(2)}\right)^{-\frac{1}{2}} y\right\|^2$$

Hence, we have that

$$\begin{aligned}\mathrm{Tr}(\Delta^{(k)}B^{(2)}\Delta^{(k)}) &\le O\left(\left\|B^{(2)}\right\|_F^{\frac{1}{2}}\right)\sqrt{\mathbb{E}_{y\sim\tilde{p}}\left\|\left(B^{(2)}\right)^{\frac{1}{2}}\Delta^{(k)}\left(B^{(2)}\right)^{-\frac{1}{2}} y\right\|^2}\\ &= O\left(\left\|B^{(2)}\right\|_F^{\frac{1}{2}}\right)\sqrt{\mathbb{E}_{y\sim\tilde{p}}\mathrm{Tr}\left(B^{(2)}\right)^{-\frac{1}{2}}\Delta^{(k)}B^{(2)}\Delta^{(k)}\left(B^{(2)}\right)^{-\frac{1}{2}} yy^T}\\ &= O\left(\left\|B^{(2)}\right\|_F^{\frac{1}{2}}\right)\sqrt{\mathrm{Tr}\Delta^{(k)}B^{(2)}\Delta^{(k)}}.\end{aligned}$$

Hence, we have that

$$\mathrm{Tr}(\Delta^{(k)}B^{(2)}\Delta^{(k)}) \le \left\|B^{(2)}\right\|_F.$$

Putting it into (4.1) gives that

$$\left|\mathbb{E}_{x,y\sim p} x^T B^{(1)} y\cdot x^T B^{(2)} y\cdot x^T B^{(3)} y\right| \le O(1)\cdot \mathrm{Tr}B^{(1)}\cdot\left\|B^{(2)}\right\|_F\cdot\left\|B^{(3)}\right\|_{\mathrm{op}}.$$

□

Lemma 23. *For $q=2$, let $u_t = u(A_t)$, we have that*

$$du_t = \alpha_t^T dW_t + \beta_t dt$$

with

$$\left\|\alpha_t\right\|_2 \le O(u_t^{\frac{3}{2}}) \quad \textit{and} \quad \beta_t \le O(u_t^3)\sqrt{\Phi}.$$

Proof. For α_t, we use Lemma 21 and get that

$$\left\|\mathbb{E}_{x\sim\tilde{p}_t}x^T(u_tI-A_t)^{-(q+1)}x\cdot x\right\|_2 = O(1)\left\|A_t\right\|_{\mathrm{op}}^{1/2}\mathrm{Tr}(A_t(u_tI-A_t)^{-(q+1)})$$
$$\leq O(u_t^{\frac{3}{2}})\kappa_t.$$

For β_t, we bound each term separately. For the first term, Lemma 22 shows that

$$\begin{aligned}
&\mathbb{E}_{x,y\sim\tilde{p}_t}x^T(u_tI-A_t)^{-1}y\cdot x^T(u_tI-A_t)^{-(q+1)}y\cdot x^Ty\\
\leq&O(1)\cdot\mathrm{Tr}(A_t^{\frac{1}{2}}(u_tI-A_t)^{-(q+1)}A_t^{\frac{1}{2}})\cdot\sqrt{\mathrm{Tr}(A_t^{\frac{1}{2}}(u_tI-A_t)^{-1}A_t^{\frac{1}{2}})^2}\cdot\left\|A_t\right\|_{\mathrm{op}}\\
\leq&O(1)\cdot\left\|A_t\right\|_{\mathrm{op}}^2\cdot\kappa_t\cdot\sqrt{\mathrm{Tr}(A_t(u_tI-A_t)^{-1}A_t(u_tI-A_t)^{-1})}\\
\leq&O(u_t^3)\cdot\kappa_t\cdot\sqrt{\mathrm{Tr}(u_tI-A_t)^{-2}}.
\end{aligned}$$

For the second term, we use Lemma 21 and get that

$$\begin{aligned}
&\mathbb{E}_{x,y\sim\tilde{p}_t}x^T(u_tI-A_t)^{-(q+1)}x\cdot y^T(u_tI-A_t)^{-(q+2)}y\cdot x^Ty\\
\leq&\left\|\mathbb{E}_{x\sim\tilde{p}_t}x^T(u_tI-A_t)^{-(q+1)}x\cdot x\right\|_2\left\|\mathbb{E}_{y\sim\tilde{p}_t}y^T(u_tI-A_t)^{-(q+2)}y\cdot y\right\|_2\\
\leq&O(1)\left\|A_t\right\|_{\mathrm{op}}^{1/2}\mathrm{Tr}(A_t(u_tI-A_t)^{-(q+1)})\cdot\left\|A_t\right\|_{\mathrm{op}}^{1/2}\mathrm{Tr}(A_t(u_tI-A_t)^{-(q+2)})\\
=&O(u_t^3)\cdot\kappa_t\cdot\mathrm{Tr}((u_tI-A_t)^{-(q+2)}).
\end{aligned}$$

For the third term, the same calculation shows that

$$\begin{aligned}
&\mathrm{Tr}((u_tI-A_t)^{-(q+2)})\cdot\mathbb{E}_{x,y\sim\tilde{p}_t}x^T(u_tI-A_t)^{-(q+1)}x\cdot y^T(u_tI-A_t)^{-(q+1)}y\cdot x^Ty\\
\leq&O(u_t^3)\cdot\mathrm{Tr}((u_tI-A_t)^{-(q+2)})\cdot\kappa_t^2.
\end{aligned}$$

Combining all three terms, we have

$$\beta_t=O(qu_t^3)\left(\sqrt{\mathrm{Tr}(u_tI-A_t)^{-2}}+\frac{\mathrm{Tr}((u_tI-A_t)^{-(q+2)})}{\kappa_t}\right).$$

Using that $q=2$, we have that

$$\beta_t\leq O(u_t^3)\left(\sqrt{\Phi}+\frac{\mathrm{Tr}((u_tI-A_t)^{-4})}{\mathrm{Tr}((u_tI-A_t)^{-3})}\right)\leq O(u_t^3)\left(\sqrt{\Phi}+\mathrm{Tr}((u_tI-A_t)^{-3})^{\frac{1}{3}}\right)=O(u_t^3)\sqrt{\Phi}.$$

□

Now, we are ready to upper bound $\left\|A_t\right\|_{\mathrm{op}}$.

Proof of Lemma 19. Consider the potential $\Psi_t=-(u_t+1)^{-2}$. Using Lemma 23, we have that

$$\begin{aligned}
d\Psi_t&=2(u_t+1)^{-3}(\alpha_t^TdW_t+\beta_tdt)-3(u_t+1)^{-4}\left\|\alpha_t\right\|^2dt\\
&\stackrel{\mathrm{def}}{=}\gamma_t^TdW_t+\eta_tdt.
\end{aligned}$$

Note that

$$\left\|\gamma_t\right\|_2^2=\left\|2(u_t+1)^{-3}\alpha_t\right\|_2^2\leq O(1)(u_t+1)^{-6}u_t^3\leq c$$

and

$$\eta_t\leq 2(u_t+1)^{-3}O(u_t^3)\sqrt{\Phi}\leq c\sqrt{\Phi}$$

for some universal constant c.

Let Y_t be the process $dY_t=\gamma_t^TdW_t$. By Theorem 15, there exists a Wiener process $\widetilde{W}_t$ such that Y_t has the same distribution as $\widetilde{W}_{[Y]_t}$. Using the reflection principle for 1-dimensional Brownian motion, we have that

$$\mathbb{P}(\max_{t\in[0,T]}Y_t\geq\gamma)\leq\mathbb{P}(\max_{t\in[0,cT]}\widetilde{W}_t\geq\gamma)=2\mathbb{P}(\widetilde{W}_{cT}\geq\gamma)\leq 2\exp(-\frac{\gamma^2}{2cT}).$$

Therefore, we have that

$$\mathbb{P}(\max_{t\in[0,T]} \Psi_t - \Psi_0 \geq c\sqrt{\Phi}T + \gamma) \leq 2\exp(-\frac{\gamma^2}{2cT}).$$

Set $\Phi = 4n$. At $t = 0$, we have $\mathrm{Tr}(u_0 I - I)^{-2} = \frac{n}{4}$. Therefore, $u_0 = \frac{3}{2}$ and $\Psi_0 = -\frac{4}{25}$. Using the assumptions that $T \leq \frac{1}{25c\sqrt{\Phi}}$, we have that

$$\mathbb{P}(\max_{t\in[0,T]} \left(-(u_t+1)^{-2}\right) \geq -\frac{3}{25} + \gamma) \leq 2\exp(-\frac{\gamma^2}{2cT}).$$

The result follows from setting $\gamma = \frac{1}{120}$. □

4.2 Calculus with the BSS potential

Here we prove Lemma 20.

Lemma 24. *We have that*

$$Du(X)[H] = \frac{\mathrm{Tr}((uI-X)^{-(q+1)}H)}{\mathrm{Tr}((uI-X)^{-(q+1)})}$$

and

$$\begin{aligned} D^2u(X)[H_1,H_2] =& \frac{\sum_{k=1}^{q+1}\mathrm{Tr}((uI-X)^{-k}H_1(uI-X)^{-(q+2-k)}H_2)}{\mathrm{Tr}((uI-X)^{-(q+1)})} \\ &-(q+1)\frac{\mathrm{Tr}((uI-X)^{-(q+1)}H_1)\mathrm{Tr}((uI-X)^{-(q+2)}H_2)}{\mathrm{Tr}((uI-X)^{-(q+1)})^2} \\ &-(q+1)\frac{\mathrm{Tr}((uI-X)^{-(q+1)}H_2)\mathrm{Tr}((uI-X)^{-(q+2)}H_1)}{\mathrm{Tr}((uI-X)^{-(q+1)})^2} \\ &+(q+1)\frac{\mathrm{Tr}((uI-X)^{-(q+1)}H_1)\mathrm{Tr}((uI-X)^{-(q+1)}H_2)\mathrm{Tr}((uI-X)^{-(q+2)})}{\mathrm{Tr}((uI-X)^{-(q+1)})^3}. \end{aligned}$$

Proof. Taking derivative of the equation (2.3), we have

$$-q\mathrm{Tr}((uI-X)^{-(q+1)})\cdot Du(X)[H] + q\mathrm{Tr}((uI-X)^{-(q+1)}H) = 0.$$

Therefore,

$$Du(X)[H] = \frac{\mathrm{Tr}((uI-X)^{-(q+1)}H)}{\mathrm{Tr}((uI-X)^{-(q+1)})}.$$

Taking derivative again on both sides,

$$\begin{aligned} D^2u(X)[H_1,H_2] =& \frac{\sum_{k=1}^{q+1}\mathrm{Tr}((uI-X)^{-k}H_1(uI-X)^{-(q+2-k)}H_2)}{\mathrm{Tr}((uI-X)^{-(q+1)})} \\ &-\frac{(q+1)\mathrm{Tr}((uI-X)^{-(q+1)}H_1)\mathrm{Tr}((uI-X)^{-(q+2)}H_2)}{\mathrm{Tr}((uI-X)^{-(q+1)})^2} \\ &-(q+1)\frac{\mathrm{Tr}((uI-X)^{-(q+2)}H_1)}{\mathrm{Tr}((uI-X)^{-(q+1)})}Du(X)[H_2] \\ &+(q+1)\frac{\mathrm{Tr}((uI-X)^{-(q+1)}H_1)\mathrm{Tr}((uI-X)^{-(q+2)})}{\mathrm{Tr}((uI-X)^{-(q+1)})^2}Du(X)[H_2]. \end{aligned}$$

Substituting the formula of $Du(X)[H_2]$ and organizing the term, we have the result. □

To simplify the first term in the Hessian, we need the following Lemma:

Lemma 25 ([7])**.** *For any positive definite matrix A, symmetric matrix Δ and any $\alpha, \beta \geq 0$, we have that*

$$\mathrm{Tr}(A^\alpha \Delta A^\beta \Delta) \leq \mathrm{Tr}(A^{\alpha+\beta}\Delta^2).$$

Proof. Without loss of generality, we can assume A is diagonal by rotating the space. Hence, we have that

$$\begin{aligned}
\mathrm{Tr}(A^{\alpha}\Delta A^{\beta}\Delta) &= \sum_{i,j} A_{ii}^{\alpha} A_{jj}^{\beta} \Delta_{ij}^{2} \\
&\leq \sum_{i,j}\left(\frac{\alpha}{\alpha+\beta} A_{ii}^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} A_{jj}^{\alpha+\beta}\right)\Delta_{ij}^{2} \\
&= \frac{\alpha}{\alpha+\beta}\sum_{i,j} A_{ii}^{\alpha+\beta}\Delta_{ij}^{2} + \frac{\beta}{\alpha+\beta}\sum_{i,j} A_{jj}^{\alpha+\beta}\Delta_{ij}^{2} \\
&= \mathrm{Tr}(A^{\alpha+\beta}\Delta^{2}).
\end{aligned}$$

□

Now, we are already to upper bound $du(A_t)$.

Proof of Lemma 20. Using Lemma 24 and Itô's formula, we have that

$$\begin{aligned}
du(A_t) =& \frac{\mathrm{Tr}((uI-A_t)^{-(q+1)}dA_t)}{\mathrm{Tr}((uI-A_t)^{-(q+1)})} \\
&+ \frac{1}{2}\sum_{ijkl} \frac{\sum_{k=1}^{q+1}\mathrm{Tr}((uI-A_t)^{-k} e_{ij} (uI-A_t)^{-(q+2-k)} e_{kl})}{\mathrm{Tr}((uI-A_t)^{-(q+1)})} d[A_{ij}, A_{kl}]_t \\
&- \frac{q+1}{2}\sum_{ijkl} \frac{\mathrm{Tr}((uI-A_t)^{-(q+1)}e_{ij})\mathrm{Tr}((uI-A_t)^{-(q+2)}e_{kl})}{\mathrm{Tr}((uI-A_t)^{-(q+1)})^2} d[A_{ij}, A_{kl}]_t \\
&- \frac{q+1}{2}\sum_{ijkl} \frac{\mathrm{Tr}((uI-A_t)^{-(q+1)}e_{kl})\mathrm{Tr}((uI-A_t)^{-(q+2)}e_{ij})}{\mathrm{Tr}((uI-A_t)^{-(q+1)})^2} d[A_{ij}, A_{kl}]_t \\
&+ \frac{q+1}{2}\sum_{ijkl} \frac{\mathrm{Tr}((uI-A_t)^{-(q+1)}e_{ij})\mathrm{Tr}((uI-A_t)^{-(q+1)}e_{kl})\mathrm{Tr}((uI-A_t)^{-(q+2)})}{\mathrm{Tr}((uI-A_t)^{-(q+1)})^3} d[A_{ij}, A_{kl}]_t.
\end{aligned}$$

For brevity, we let $\tilde{p}_t$ be the translation of p_t that has mean 0, i.e. $\tilde{p}_t(x) = p_t(x+\mu_t)$. Let $\Delta^{(z)} = \mathbb{E}_{x\sim\tilde{p}_t} x x^T x_z$. Then, Lemma 12 shows that $dA_t = \sum_z \Delta_z dW_{t,z} - A_t^2 dt$ where $W_{t,z}$ is the z^{th} coordinate of W_t. Therefore,

$$d[A_{ij}, A_{kl}]_t = \sum_z \Delta_{ij}^{(z)}\Delta_{kl}^{(z)} dt. \tag{4.2}$$

Using the formula for dA_t (2.2) and $d[A_{ij}, A_{kl}]_t$ (4.2), we have that

$$\begin{aligned}
du(A_t) =& \frac{\mathrm{Tr}\left((uI-A_t)^{-(q+1)}\left(\mathbb{E}_{x\sim\tilde{p}_t} x x^T x^T dW_t - A_t^2 dt\right)\right)}{\mathrm{Tr}((uI-A_t)^{-(q+1)})} \\
&+ \frac{1}{2}\sum_{z} \frac{\sum_{k=1}^{q+1}\mathrm{Tr}((uI-A_t)^{-k}\Delta^{(z)}(uI-A_t)^{-(q+2-k)}\Delta^{(z)})}{\mathrm{Tr}((uI-A_t)^{-(q+1)})} dt \\
&- (q+1)\sum_{z} \frac{\mathrm{Tr}((uI-A_t)^{-(q+1)}\Delta^{(z)})\mathrm{Tr}((uI-A_t)^{-(q+2)}\Delta^{(z)})}{\mathrm{Tr}((uI-A_t)^{-(q+1)})^2} dt \\
&+ \frac{q+1}{2}\sum_{z} \frac{\mathrm{Tr}((uI-A_t)^{-(q+1)}\Delta^{(z)})\mathrm{Tr}((uI-A_t)^{-(q+1)}\Delta^{(z)})\mathrm{Tr}((uI-A_t)^{-(q+2)})}{\mathrm{Tr}((uI-A_t)^{-(q+1)})^3} dt.
\end{aligned}$$

Using Lemma 25,

$$\mathrm{Tr}((uI-A_t)^{-k}\Delta^{(z)}(uI-A_t)^{-(q+2-k)}\Delta^{(z)}) \leq \mathrm{Tr}((uI-A_t)^{-1}\Delta^{(z)}(uI-A_t)^{-(q+1)}\Delta^{(z)})$$

for all $1 \leq k \leq q+1$.

Let $\kappa_t = \mathrm{Tr}((uI - A_t)^{-(q+1)})$, then

$$\begin{aligned}du(A_t) \leq & \frac{1}{\kappa_t}\mathbb{E}_{x\sim\tilde{p}_t} x^T(uI - A_t)^{-(q+1)}xx^T dW_t \\ & + \frac{q+1}{2\kappa_t}\mathbb{E}_{x,y\sim\tilde{p}_t}\sum_z \mathrm{Tr}((uI - A_t)^{-1}xx^T x_z (uI - A_t)^{-(q+1)}yy^T y_z)dt \\ & - \frac{q+1}{\kappa_t^2}\mathbb{E}_{x,y\sim\tilde{p}_t}\sum_z \mathrm{Tr}((uI - A_t)^{-(q+1)}xx^T x_z)\mathrm{Tr}((uI - A_t)^{-(q+2)}yy^T y_z)dt \\ & + \frac{q+1}{2\kappa_t^3}\mathbb{E}_{x,y\sim\tilde{p}_t}\sum_z \mathrm{Tr}((uI - A_t)^{-(q+1)}xx^T x_z)\mathrm{Tr}((uI - A_t)^{-(q+1)}yy^T y_z)\mathrm{Tr}((uI - A_t)^{-(q+2)})dt.\end{aligned}$$

Rearranging the terms, we have the result. □

4.3 Bounding the size of any initial set

Fix any set $E \subset \mathbb{R}^n$ and define $g_t = p_t(E)$.

Lemma 26. *The random variable g_t is a martingale satisfying*

$$d[g_t]_t \leq D^2 g_t^2 dt$$

and

$$d[g_t]_t \leq 30 \left\|A_t\right\|_{\mathrm{op}} \cdot g_t^2 \log^2\left(\frac{1}{g_t}\right) dt.$$

Proof. Note that

$$dg_t = \left\langle \int_E (x - \mu_t)p_t(x)dx, dW_t \right\rangle.$$

Therefore, we have that

$$d[g_t]_t = \left\| \int_E (x - \mu_t)p_t(x)dx \right\|_2^2 dt.$$

We bound the norm in two different ways. On one hand, we note that

$$\left\| \int_E (x - \mu_t)p_t(x)dx \right\|_2 \leq \sup \left\|x - \mu_t\right\|_2 \left(\int_E p_t(x)dx \right) = \sup \left\|x - \mu_t\right\|_2 g_t \leq D \cdot g_t \tag{4.3}$$

On the other hand, for any $k \geq 1$, we have that

$$\begin{aligned}\left\| \int_E (x - \mu_t)p_t(x)dx \right\|_2 &= \max_{\|\zeta\|_2=1} \int_E (x - \mu_t)^T\zeta \cdot p_t(x)dx \\ &\leq \max_{\|\zeta\|_2=1} \left(\int_E \left|(x - \mu_t)^T\zeta\right|^k \cdot p_t(x)dx \right)^{\frac{1}{k}} \left(\int_E p_t(x)dx \right)^{1-\frac{1}{k}} \\ &\leq 2k \left\|A_t\right\|_{\mathrm{op}}^{1/2} \cdot g_t^{1-\frac{1}{k}} \end{aligned} \tag{4.4}$$

where we used Lemma 16 at the end. Setting $k = \log(\frac{1}{g_t})$, we have the result. □

Using this, we can bound how fast $\log \frac{1}{g_t}$ changes.

Lemma 27. *For any $T \geq 0$ and $\gamma \geq 0$, we have that*

$$\mathbb{P}\left(\text{for all } 0 \leq t \leq T: \log\frac{1}{g_0} + \frac{1}{2}D^2 t + \gamma \geq \log\frac{1}{g_t} \geq \log\frac{1}{g_0} - \gamma \right) \geq 1 - 4\exp(-\frac{\gamma^2}{2TD^2}).$$

Proof. Since $dg_t = g_t \alpha_t^T dW_t$ for some $\|\alpha_t\|_2 \leq D$ (from (4.3) in Lemma 26), using Itô's formula (Lemma (13)) we have that

$$\begin{aligned} d\log\frac{1}{g_t} &= -\frac{dg_t}{g_t} + \frac{1}{2}\frac{d[g_t]_t}{g_t^2} \\ &= -\alpha_t^T dW_t + \frac{1}{2}\|\alpha_t\|^2 dt. \end{aligned}$$

Let Y_t be the process $dY_t = \alpha_t^T dW_t$. By Theorem 15 and the reflection principle, we have that

$$\mathbb{P}(\max_{t\in[0,T]} |Y_t| \geq \gamma) \leq 4\exp(-\frac{\gamma^2}{2TD^2}).$$

□

Now, we bound $\mathbb{E} g_t\sqrt{\log\frac{1}{g_t}}$. This is the main result of this section.

Lemma 28. *Assume that $n \geq 10$. There is some universal constant $c \geq 0$ such that for any measurable subset E such that $p_0(E) \leq c$ and any T such that*

$$0 \leq T \leq c \cdot \max\left(\frac{1}{D^2}\log\frac{1}{p_0(E)}, \frac{1}{\log\frac{1}{p_0(E)} + D}\right),$$

we have that

$$\mathbb{E}\left(p_T(E)\sqrt{\log\frac{1}{p_T(E)}} 1_{p_T(E)\leq\frac{1}{2}}\right) \geq \frac{1}{5}p_0(E)\sqrt{\log\frac{1}{p_0(E)}}.$$

Proof. Fix any set $E \subset \mathbb{R}^n$ and define $g_t = p_t(E)$.

Case 1) $T \leq \frac{1}{8}D^{-2}\log\frac{1}{g_0}$. Lemma 27 shows that

$$\mathbb{P}(\log\frac{1}{g_t} \geq \frac{1}{4}\log\frac{1}{g_0} \text{ for all } 0 \leq t \leq T) \geq 1 - 4g_0^2.$$

Using $g_0 \leq \frac{1}{16}$, we have that

$$\mathbb{E}\left(g_T\sqrt{\log\frac{1}{g_T}} 1_{g_T\leq\frac{1}{2}}\right) \geq \mathbb{E}\left(g_T\sqrt{\log\frac{1}{g_T}} 1_{\log\frac{1}{g_T}\geq\frac{1}{4}\log\frac{1}{g_0}}\right) \geq \mathbb{E}\left(g_T 1_{\log\frac{1}{g_T}\geq\frac{1}{4}\log\frac{1}{g_0}}\right)\sqrt{\frac{1}{4}\log\frac{1}{g_0}}.$$

Since $\mathbb{E}g_T = g_0$ and $g_T \leq 1$, we have that

$$\mathbb{E}\left(g_T 1_{\log\frac{1}{g_T}\geq\frac{1}{4}\log\frac{1}{g_0}}\right) = g_0 - \mathbb{E}\left(g_T 1_{\log\frac{1}{g_T}<\frac{1}{4}\log\frac{1}{g_0}}\right) \geq g_0 - 4g_0^2 \geq \frac{1}{2}g_0.$$

Therefore,

$$\mathbb{E}\left(g_T\sqrt{\log\frac{1}{g_T}} 1_{g_T\leq\frac{1}{2}}\right) \geq \frac{1}{4}g_0\sqrt{\log\frac{1}{g_0}}.$$

Case 2) $T \geq \frac{1}{8}D^{-2}\log\frac{1}{g_0}$. Now, we assume that $T \leq \frac{1}{2c(D+\log\frac{1}{g_0})}$ where $c \geq 1$ is the universal constant appears in Lemma 19. Note that

$$dg_t\sqrt{\log\frac{e}{g_t}} = \frac{2\log\frac{e}{g_t} - 1}{2\sqrt{\log\frac{e}{g_t}}}dg_t - \frac{2\log\frac{e}{g_t} + 1}{8g_t\log^{\frac{3}{2}}\frac{e}{g_t}}d[g_t]_t.$$

Since $dg_t = g_t\log\frac{e}{g_t}\alpha_t^T dW_t$ for some $\|\alpha_t\|_2 \leq \sqrt{30}\|A_t\|_{\text{op}}^{1/2}$ (from (4.4) in Lemma 26),

$$dg_t\sqrt{\log\frac{e}{g_t}} = \frac{1}{2}g_t\sqrt{\log\frac{e}{g_t}}(2\log\frac{e}{g_t} - 1)\alpha_t^T dW_t - \frac{1}{8}g_t\sqrt{\log\frac{e}{g_t}}(2\log\frac{e}{g_t} + 1)\|\alpha_t\|_2^2 dt.$$

For any $s \ge s' \ge 0$, we have that

$$\begin{aligned}\mathbb{E} g_s \sqrt{\log \frac{e}{g_s}} &= g_{s'} \sqrt{\log \frac{e}{g_{s'}}} - \frac{1}{8} \int_{s'}^{s} \mathbb{E}\left(g_t \sqrt{\log \frac{e}{g_t}} (2 \log \frac{e}{g_t} + 1) \left\| \alpha_t \right\|_2^2 \right) dt \\ &\ge g_{s'} \sqrt{\log \frac{e}{g_{s'}}} - 12 \int_{s'}^{s} \mathbb{E}\left(\left\| A_t \right\|_{\mathrm{op}} g_t \log^{\frac{3}{2}} \frac{e}{g_t} \right) dt. \end{aligned} \tag{4.5}$$

Using $T \ge \frac{1}{8} D^{-2} \log \frac{1}{g_0}$, Lemma 27 shows that

$$\mathbb{P}(15 D^2 T \ge \max_{0 \le t \le T} \log \frac{1}{g_t}) \ge 1 - 4 g_0^2. \tag{4.6}$$

Now, using $T \le \frac{1}{2c(\sqrt{n} + \log \frac{1}{g_0})}$, Lemma 19 shows that

$$\mathbb{P}(\max_{t \in [0,T]} \left\| A_t \right\|_{\mathrm{op}} \ge 2) \le 2 g_0^2. \tag{4.7}$$

Let E be the event that both $\max_{0 \le t \le T} \log \frac{1}{g_t} \le 15 D^2 T$ and $\max_{t \in [0,T]} \left\| A_t \right\|_{\mathrm{op}} \le 2$. Then, combining (4.6) and (4.7), we have that

$$\begin{aligned}\mathbb{E}\left(\left\| A_t \right\|_{\mathrm{op}} g_t \log^{\frac{3}{2}} \frac{e}{g_t} \right) &\le 2(1 + 15 D^2 T) \mathbb{E}\left(g_t \log^{\frac{1}{2}} \frac{e}{g_t} 1_E \right) + \mathbb{E}\left(\left\| A_t \right\| g_t \log^{\frac{3}{2}} \frac{e}{g_t} 1_{E^c} \right) \\ &\le 2(1 + 15 D^2 T) \mathbb{E}\left(g_t \log^{\frac{1}{2}} \frac{e}{g_t} \right) + 12 \sqrt{D} g_0^2. \end{aligned}$$

where we used that $\left\| A_t \right\|_{\mathrm{op}} \le \sqrt{D}$ a.s. and $g_t \log^{\frac{3}{2}} \frac{e}{g_t} \le 2$ and $\mathbb{P}(E^c) \le 6 g_0^2$. Putting it into (4.5), for any $s' \le s \le T$, we have that

$$\mathbb{E} g_s \sqrt{\log \frac{e}{g_s}} \ge g_{s'} \sqrt{\log \frac{e}{g_{s'}}} - 24(1 + 15 D^2 T) \int_{s'}^{s} \mathbb{E} g_t \sqrt{\log \frac{e}{g_t}} dt - 144 \sqrt{D} g_0^2 T.$$

Let s^* be the $s \in [0,T]$ that maximizes $\mathbb{E} g_s \sqrt{\log \frac{e}{g_s}}$ and $f_t = \mathbb{E} g_t \sqrt{\log \frac{e}{g_t}}$. For all $T \ge s \ge s^*$, we have that

$$\begin{aligned} f_s &\ge f_{s^*} - 24(1 + 15 D^2 T) \int_{s'}^{s} f_t dt - 144 \sqrt{D} g_0^2 T \\ &\ge f_{s^*} - 24(T + 15 D^2 T^2) f_{s^*} - 144 \sqrt{D} g_0^2 T \\ &\ge \frac{4}{5} f_{s^*} \ge \frac{4}{5} f_0 \end{aligned}$$

where we used that $T \le \frac{1}{10^5 D} \le \frac{1}{10^5}$ at the end. Therefore, we have

$$\mathbb{E} g_T \sqrt{\log \frac{e}{g_T}} \ge \frac{4}{5} g_0 \sqrt{\log \frac{e}{g_0}}.$$

Now, we note that

$$\begin{aligned}\mathbb{E}\left(g_T \sqrt{\log \frac{1}{g_T}} 1_{g_T \le \frac{1}{2}} \right) &\ge \frac{1}{2} \mathbb{E}\left(g_T \sqrt{\log \frac{e}{g_T}} 1_{g_T \le \frac{1}{2}} \right) \ge \frac{2}{5} g_0 \sqrt{\log \frac{e}{g_0}} - \frac{\sqrt{\log 2e}}{2} g_0 \\ &\ge \frac{1}{5} g_0 \sqrt{\log \frac{e}{g_0}} \end{aligned}$$

where we used $g_0 \le \frac{1}{e^{10}}$ at the end. $\square$

4.4 Gaussian Case

The next theorem can be found in [15, Thm 1.1]. We give another proof for completeness.

Theorem 29. *Let* $h(x) = f(x)e^{-\frac{t}{2}\|x\|^2} / \int f(y)e^{-\frac{t}{2}\|y\|^2} dy$ *where* $f : \mathbb{R}^n \to \mathbb{R}_+$ *is an integrable logconcave function and* $t \geq 0$. *Let* $p(S) = \int_S h(x)dx$. *For any* $p(S) \leq \frac{1}{2}$, *we have*

$$p(\partial S) = \Omega\left(\sqrt{t}\right) \cdot p(S)\sqrt{\log \frac{1}{p(S)}}.$$

Proof. Let $g = p(S)$. Then, the desired statement can be written as

$$\int_S h(x)\, dx = g \int_{\mathbb{R}^n} h(x) \implies \int_{\partial S} h(x)\, dx \geq c\sqrt{t} g \sqrt{\log(\frac{1}{g})} \int_{\mathbb{R}^n} h(x)$$

for some constant c. By the localization lemma [11], if there is a counterexample, there is a counterexample in one-dimension where h is of the form $h(x) = Ce^{-\gamma x - t\frac{x^2}{2}}$ restricted to some interval on the real line. Without loss of generality, we can assume that S is a single interval, otherwise, any interval has smaller boundary measure and smaller volume. By rescaling, flipping and shifting the function h, we can assume that $t = 1$ and that $h(x) = e^{-\frac{1}{2}x^2} 1_{[a,b]}$ and that $S = [y, b]$ for some $a < y < b$.

It remains to show that for $g \leq \frac{1}{2}$,

$$\frac{\int_y^b e^{-\frac{x^2}{2}}\, dx}{\int_a^b e^{-\frac{x^2}{2}}\, dx} = g \implies \frac{e^{-\frac{y^2}{2}}}{\int_a^b e^{-\frac{x^2}{2}}\, dx} \gtrsim g\sqrt{\log \frac{1}{g}}.$$

There are three cases: $y \geq 1$, $y \leq -1$ and $-1 \leq y \leq -1$. Let $A = \int_a^b e^{-\frac{x^2}{2}}\, dx$. In the first case $y \geq 1$, we note that the integral

$$g \cdot A = \int_y^b e^{-\frac{x^2}{2}}\, dx \leq \int_y^\infty e^{-\frac{x^2}{2}}\, dx \lesssim e^{-\frac{y^2}{2}}/y. \tag{4.8}$$

Rearrange the terms, we have that

$$\frac{e^{-\frac{y^2}{2}}}{\int_a^b e^{-\frac{x^2}{2}}\, dx} \gtrsim g \cdot y.$$

Using (4.8), we have that $y \gtrsim \sqrt{\log \frac{1}{g \cdot A}}$ and hence

$$\frac{e^{-\frac{y^2}{2}}}{\int_a^b e^{-\frac{x^2}{2}}\, dx} \gtrsim g \cdot \sqrt{\log \frac{1}{g \cdot A}} \gtrsim g \cdot \sqrt{\log \frac{1}{g}}$$

where we used that $A = \int_a^b e^{-\frac{x^2}{2}}\, dx \leq \sqrt{2\pi}$ and $g \leq \frac{1}{2}$.

For the second case $y \leq -1$, since $g \leq \frac{1}{2}$, we have that

$$\frac{e^{-\frac{y^2}{2}}}{\int_a^b e^{-\frac{x^2}{2}}\, dx} \geq \frac{2e^{-\frac{y^2}{2}}}{\int_a^y e^{-\frac{x^2}{2}}\, dx} \geq \frac{e^{-\frac{y^2}{2}}}{\int_{-\infty}^y e^{-\frac{x^2}{2}}\, dx} \gtrsim |y| \geq 1$$

where we used that $\int_{-\infty}^y e^{-\frac{x^2}{2}}\, dx = \frac{O(1)}{|y|} e^{-\frac{y^2}{2}}$ at the end. This proves the second case because $g\sqrt{\log \frac{1}{g}} \lesssim 1$.

For the last case $|y| \leq 1$, we note that

$$\frac{e^{-\frac{y^2}{2}}}{\int_a^b e^{-\frac{x^2}{2}}\, dx} \geq \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} \gtrsim 1.$$

$\square$

4.5 Proof of main theorem

We can now prove a bound on the isoperimetric constant.

Proof of Theorem. 18. By Lemma 12, p_t is a martingale and therefore

$$p(\partial E) = p_0(\partial E) = \mathbb{E} p_T(\partial E).$$

Next, by the definition of p_T (2.1), we have that $p_T(x) \propto e^{c_T^T x - \frac{T}{2}\|x\|^2} p(x)$ and Theorem 29 shows that if $p_t(E) \leq \frac{1}{2}$, we have that

$$p_T(\partial E) \gtrsim \sqrt{T} \cdot p_t(E) \sqrt{\log \frac{1}{p_t(E)}}.$$

Hence, we have that

$$p(\partial E) \gtrsim \sqrt{T} \cdot \mathbb{E} \left(p_t(E) \sqrt{\log \frac{1}{p_t(E)}} \cdot 1_{p_t(E) \leq \frac{1}{2}} \right).$$

Lemma 28 shows that if

$$T \leq c \cdot \max \left(\frac{1}{D^2} \log \frac{1}{p_0(E)}, \frac{1}{\log \frac{1}{p_0(E)} + D} \right)$$

and $p_0(E) \leq c$ for some small enough constant c, we have

$$\begin{aligned}
p(\partial E) &\gtrsim \sqrt{T} \cdot p_0(E) \sqrt{\log \frac{1}{p_0(E)}} \\
&\gtrsim \left(\frac{\sqrt{\log \frac{1}{p_0(E)}}}{D} + \frac{1}{\sqrt{\log \frac{1}{p_0(E)} + D}} \right) p_0(E) \sqrt{\log \frac{1}{p_0(E)}} \\
&\gtrsim \left(\frac{\log \frac{1}{p_0(E)}}{D} + \sqrt{\frac{\log \frac{1}{p_0(E)}}{\log \frac{1}{p_0(E)} + D}} \right) p_0(E) \\
&\gtrsim \left(\frac{\log \frac{1}{p_0(E)}}{D} + \sqrt{\frac{\log \frac{1}{p_0(E)}}{D}} \right) p_0(E).
\end{aligned}$$

If $p_0(E) \geq c$, the bound simply follows from Theorem 2. □

5 Consequences of the improved isoperimetric inequality

In this section, we give some consequences of the improved isoperimetric inequality (Theorem 18). First we note the following Cheeger-type logarithmic isoperimetric inequality.

Theorem 30 ([14])**.** *Let μ be any absolutely continuous measure on $\mathbb{R}^n$ such that for any open subsets A of $\mathbb{R}^n$ with $\mu(A) \leq \frac{1}{2}$,*

$$\mu(\partial A) \geq \phi \cdot \mu(A) \sqrt{\log \frac{1}{\mu(A)}}.$$

Then, for any f such that $\int_{\mathbb{R}^n} f^2 d\mu = 1$, we have that

$$\int_{\mathbb{R}^n} |\nabla f|^2 \, d\mu \gtrsim \phi^2 \cdot \int_{\mathbb{R}^n} f^2 \log f^2 d\mu.$$

Applying this and Theorem 18, we have the following result.

Theorem 31. *Given an isotropic logconcave distribution p with diameter D. For any f such that $\int_{\mathbb{R}^n} f^2(x)p(x)dx = 1$, we have that*

$$\int_{\mathbb{R}^n} |\nabla f(x)|^2 p(x)dx \gtrsim \frac{1}{D} \cdot \int_{\mathbb{R}^n} f^2 \log f^2 p(x)dx.$$

Remark. In general, isotropic logconcave distributions with diameter D can have the coefficient as small as $O(\frac{1}{D})$, as we show in Section 6. Therefore, our bound is tight up to constant.

Now we prove Theorem 9, an improved concentration inequality for general Lipschitz functions and general isotropic logconcave densities.

Proof of Theorem 9. We first prove the statement for the function $g(x) = \|x\|$. Define

$$E_t = \{x \text{ such that } \|x\| \geq \text{med}_{x\sim p} \|x\| + t\}$$

and $\alpha_t \stackrel{\text{def}}{=} \log \frac{1}{p(E_t)}$. By the definition of median, we have that $\alpha_0 = \log 2$. We first give a weak estimate on how fast α_t increases. Note that

$$\frac{d\alpha_t}{dt} = -\frac{1}{p(E_t)}\frac{dp(E_t)}{dt} = \frac{p(\partial E_t)}{p(E_t)} \geq \frac{c_1}{n^{1/4}}$$

for some universal constant $c_1 > 0$ where we used that definition of $p(\partial E_t)$ to get $\frac{dp(E_t)}{dt} = -p(\partial E_t)$ and Theorem 2 at the end. Therefore,

$$\alpha_{t+s} \geq \alpha_t + c_1 \cdot sn^{-1/4} \tag{5.1}$$

for all $t, s \geq 0$. To improve on this bound, we consider the distribution p_t defined by truncating the distribution p to the set $\|x\| \leq \text{med}_{x\sim p} \|x\| + t + c_2\sqrt{n}$ for some large enough constant c_2. By the estimate (5.1), we see that $p(E_t)$ only decreases by a tiny factor after truncation and hence

$$\frac{p(\partial E_t)}{p(E_t)} \geq \frac{1}{2}\frac{p_t(\partial E_t)}{p_t(E_t)}.$$

Next, we note that p_t is almost isotropic, namely its covariance matrix A_t satisfies $\frac{1}{2}I \preceq A_t \preceq 2I$ (in fact, it is exponentially close to I). Although Theorem 18 only applies to the isotropic case, but we can always renormalize the distribution p_t to isotropic and then normalize it back. Since the distribution p_t is almost isotropic, such a re-normalization does not change $p_t(\partial E)$ by more than a multiplicative constant.

Therefore, we can apply Theorem 18 and get that

$$\frac{d\alpha_t}{dt} = \frac{p(\partial E_t)}{p(E_t)} \geq \frac{1}{2}\frac{p_t(\partial E_t)}{p_t(E_t)} \gtrsim \sqrt{\frac{\log \frac{1}{p_t(E_t)}}{m+t}} \geq c_3 \cdot \sqrt{\frac{\alpha_t}{m+t}} \tag{5.2}$$

for some universal constant $c_3 > 0$ where $m = \text{med}_{x\sim p} \|x\| + c_2\sqrt{n}$. Note that

$$\frac{d\sqrt{\alpha_t}}{dt} \geq \frac{c_3}{2} \cdot \sqrt{\frac{1}{m+t}}$$

and hence

$$\sqrt{\alpha_t} - \sqrt{\alpha_0} \geq \frac{c_3}{2} \int_0^t \sqrt{\frac{1}{m+s}} ds = c_3 \cdot \left(\sqrt{m+t} - \sqrt{m}\right).$$

For $t \leq m$, we have that

$$\sqrt{\alpha_t} - \sqrt{\alpha_0} \geq c_3 \cdot \left(\sqrt{m} + \sqrt{m}\frac{t}{3m} - \sqrt{m}\right) \geq \frac{c_3 \cdot t}{3\sqrt{m}}.$$

and hence $\alpha_t \geq \frac{c_3^2}{9m} t^2$.

For $t \geq m$, we note that $\sqrt{1+x} - 1 \geq \frac{1}{3}\sqrt{x}$ for all $x \geq 1$. Therefore,

$$\sqrt{m+t} - \sqrt{m} = \sqrt{m}\left(\sqrt{1+\frac{t}{m}} - 1\right) \geq \frac{1}{3}\sqrt{m}\sqrt{\frac{t}{m}} = \frac{1}{3}\sqrt{t}.$$

Hence, (5.2) shows that $\sqrt{\alpha_t} - \sqrt{\alpha_0} \geq \frac{c_3}{3}\sqrt{t}$.

Combining both cases, we have that $\alpha_t \gtrsim \min\left(\frac{t^2}{\sqrt{n}}, t\right) \gtrsim \frac{t^2}{t+\sqrt{n}}$. Hence,

$$\mathbb{P}_{x\sim p}\left(\|x\| - \mathrm{med}_{x\sim p}\|x\| \geq t\right) \leq \exp\left(-\Omega(1)\cdot\frac{t^2}{t+\sqrt{n}}\right). \tag{5.3}$$

By the same proof, we have that

$$\mathbb{P}_{x\sim p}\left(\|x\| - \mathrm{med}_{x\sim p}\|x\| \leq -t\right) \leq \exp\left(-\Omega(1)\cdot\frac{t^2}{t+\sqrt{n}}\right).$$

This completes the proof for the Euclidean norm.For a general 1-Lipschitz function g, we define

$$E_t = \{x \text{ such that } g(x) \geq m_g + t\}$$

where $m_g = \mathrm{med}_{x\sim p} g(x)$. Again, we define $\alpha_t \stackrel{\text{def}}{=} \log\frac{1}{p(E_t)}$ and we have $\alpha_0 = \log 2$. To compute $\frac{p(\partial E_t)}{p(E_t)}$, we consider the restriction of p to a large ball. Let $\zeta \geq 0$ to be chosen later and $B_{t,\zeta}$ be the ball centered at 0 with radius $c_4 \cdot (\sqrt{n} + \alpha_t) + \zeta$ where c_4 is a constant. Let $p_{t,\zeta}$ be the distribution defined by

$$p_{t,\zeta}(A) = \frac{p(A\cap B_{t,\zeta})}{p(B_{t,\zeta})}.$$

Choosing a large constant c_4 and using (5.3), we have that $p(B_{t,\zeta}^c) \leq \frac{p(E_t)}{100n}$ and that $p_{t,\zeta}$ is almost isotropic. Since $p(B_{t,\zeta}^c)$ is so small, we have that $p_{t,\zeta}(E_t) \geq \frac{1}{2}p(E_t)$ and that

$$p_{t,\zeta}(\partial E_t) = \frac{p(\partial(E_t\cap B_{t,\zeta}))}{p(B_{t,\zeta})} \leq 2p(\partial E_t) + 2p(\partial B_{t,\zeta}).$$

Hence,

$$\frac{p(\partial E_t)}{p(E_t)} \geq \frac{1}{4}\frac{p_{t,\zeta}(\partial E_t)}{p_{t,\zeta}(E_t)} - \frac{p_{t,\zeta}(\partial B_{t,\zeta})}{p(E_t)}. \tag{5.4}$$

Since $p_{t,\zeta}$ is almost isotropic and has support of diameter $O(\sqrt{n} + \alpha_t + \zeta)$, Theorem 18 gives that

$$\frac{p_{t,\zeta}(\partial E_t)}{p_{t,\zeta}(E_t)} \gtrsim \sqrt{\frac{\log\frac{1}{p_{t,\zeta}(E_t)}}{\sqrt{n}+\alpha_t+\zeta}} \geq c_5\sqrt{\frac{\alpha_t}{\sqrt{n}+\alpha_t+\zeta}} \tag{5.5}$$

for some universal constant $0 < c_5 < 1$. To bound the second term $p_{t,\zeta}(\partial B_{t,\zeta})$, we note that

$$\int_0^{1/c_5^2} p_{t,\zeta}(\partial B_{t,\zeta})d\zeta \leq 2\int_0^{1/c_5^2} p(\partial B_{t,\zeta})d\zeta \leq 2p(B_{t,0}^c) \leq \frac{p(E_t)}{50n}$$

Hence, there is ζ between 0 and $1/c_5^2$ such that $p_{t,\zeta}(\partial B_{t,\zeta}) \leq \frac{p(E_t)}{50n}c_5^2$.

Using this, (5.4) and (5.5), we have that

$$\begin{aligned}\frac{p(\partial E_t)}{p(E_t)} &\geq \frac{c_5}{4}\sqrt{\frac{\alpha_t}{\sqrt{n}+\alpha_t+1/c_5^2}} - \frac{c_5^2}{50n}\\ &\geq \frac{c_5}{8}\sqrt{\frac{\alpha_t}{\sqrt{n}+\alpha_t+1/c_5^2}} + \frac{c_5}{16}\sqrt{\frac{1}{\sqrt{n}+1/c_5^2}} - \frac{c_5^2}{50n}\\ &\geq \frac{c_5}{8}\sqrt{\frac{\alpha_t}{\sqrt{n}+\alpha_t+1/c_5^2}}.\end{aligned} \tag{5.6}$$

Next, we relate $\frac{p(\partial E_t)}{p(E_t)}$ with $dp(E_t)$. Since g is 1-Lipschitz, for any x such that $\|x-y\|_2 \leq h$ and $y \in E_t$, we have that

$$g(x) \geq m_g + t - h$$

Therefore, we have $B(E_t, h) \subset E_{t-h}$ and,

$$-\frac{dp(E_t)}{dt} = \lim_{h\to 0} \frac{p\left(\{g(x) \geq \text{med}_{x\sim p} g(x) + t - h\}\right) - p(E_t)}{h}$$
$$\geq \lim_{h\to 0} \frac{p\left(B(E_t, h)\right) - p(E_t)}{h} = p(\partial E_t).$$

Using this with (5.6), we have that

$$\frac{d\alpha_t}{dt} = -\frac{1}{p(E_t)}\frac{dp(E_t)}{dt} \geq \frac{p(\partial E_t)}{p(E_t)} \geq \frac{c_5}{8}\sqrt{\frac{\alpha_t}{\sqrt{n} + \alpha_t + 1/c_5^2}}.$$

Solving this equation, we again have that $\alpha_t \gtrsim \frac{t^2}{t+\sqrt{n}}$. This proves that

$$\mathbb{P}_{x\sim p}\left(g(x) - \text{med}_{x\sim p} g(x) \geq t\right) \leq \exp\left(-\Omega(1)\cdot \frac{t^2}{t+\sqrt{n}}\right).$$

The case of $g(x) - \text{med}_{x\sim p} g(x) \leq -t$ is the same by taking the negative of g. To replace $\text{med}_{x\sim p} g(x)$ by $\mathbb{E}_{x\sim p} g(x)$, we simply use the concentration we just proved to show that $|\mathbb{E}_{x\sim p} g(x) - \text{med}_{x\sim p} g(x)| \lesssim n^{\frac{1}{4}}$. $\square$

6 Optimality of the bounds

Lemma 32. *For any $\frac{n}{2} \geq D \geq 2\sqrt{n}$, there exists an isotropic logconcave distribution with diameter D such that its log-Cheeger constant is $O(1/\sqrt{D})$ and its log-Sobolev constant is $O(1/D)$.*

In fact, we get a nearly lower tight bound on the mixing time of the ball walk, from an arbitrary start, in terms of the number of proper steps. Recall that by a proper step we mean steps where the current point changes, not counting the steps that are discarded due to the rejection probability. In our lower bound example, the local conductance, i.e., the probability of a proper step, is at least a constant everywhere and so the total number of steps in expectation is within a constant factor of the number of proper steps.

Lemma 33. *For any $\frac{n}{2} \geq D \geq 2\sqrt{n}$, there exists an isotropic convex body with diameter D such that ball walk mixes in $\widetilde{\Omega}(n^2 D)$ proper steps.*

Both theorems are based on the following cone

$$K = \left\{x:\ 0 \leq x_1 \leq n, \sum_{i=2}^{n} x_i^2 \leq \frac{1}{n}x_1^2\right\},$$

and the truncated cone

$$K_D = K \cap \left\{x:\ (x_1 - n)^2 + \sum_{i=2}^{n} x_i^2 \leq D^2\right\}.$$

Proof of Lemma 32. The convex body K_D is nearly isotropic and has diameter D. Let

$$t_0 = n - \sqrt{D^2 - n}.$$

Consider the subset $S = K \cap \{t_0 \leq x_1 \leq t_0 + 1\}$. Note that S is fully contained in K_D and that

$$p = \frac{\text{vol}(S)}{\text{vol}(K_D)} \approx \frac{\text{vol}(S)}{\text{vol}(K)}$$
$$= \left(\frac{t_0+1}{n}\right)^{n-1} - \left(\frac{t_0}{n}\right)^{n-1}$$
$$\approx e^{-D}.$$

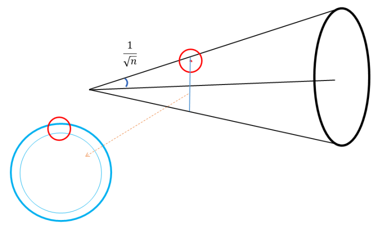

Figure 6.1: The lower bound construction

On the other hand, the expansion of S is at most 2. Therefore, the log-Cheeger constant κ of K_D must satisfy

$$\kappa\sqrt{\log\frac{1}{p}} \leq 2$$

or $\kappa = O\left(D^{-\frac{1}{2}}\right)$ as claimed. It is known that the log-Sobolev constant $\rho = \Theta(\kappa^2)$ (see e.g., [14]). This gives the second claim. □

The proof of the lower bound for the ball walk is more involved.

Proof of Lemma 33. Let the starting distribution be the uniform distribution in the set $S = K_D \cap \{t_0 \leq x_1 \leq t_0 + 1\}$. For each point x, the local conductance is $\ell(x) = \frac{\text{vol}(K_D\cap(x+\delta B))}{\text{vol}(\delta B)}$, the fraction of the δ-ball around x contained in K_D.

The distribution at any time will remain spherically symmetric at each time step. Each step along the e_1 direction is approximately $\pm\frac{\delta}{\sqrt{n}} = \pm\frac{1}{n}$. An unbiased process that moves $\pm\frac{1}{n}$ along e_1 in each step would take $\Omega(n^2D^2)$ steps to converge since the diameter is effectively nD. But there is a slight drift in the positive direction towards the base. This is because for points near the boundary of the cone a step away from the base has higher rejection probability compared to a step towards the base. We will now upper bound the drift and therefore lower bound the number of steps needed.

More precisely, let $\delta = \frac{1}{\sqrt{n}}$ and $D < 0.99n$. On any slice $A(t)$, we can identify the bias felt by each point $y \in A(t)$. It is the fraction of the δ-ball around y that is contained in K but its mirror image through the plane $x_1 = y_1$ is not contained in K. This fraction depends on the distance d of y to the boundary of $A(t)$. The points that feel a significant bias are contained in the outer annulus of thickness $O\left(\frac{\delta}{\sqrt{n}}\right) = O\left(\frac{1}{n}\right)$. To bound the fraction of the distribution in this annulus, we observe that the density within each slice is spherically symmetric. In fact, a stronger property holds.

Claim 34. The distribution at any time in any cross-section is spherically symmetric and unimodal.

Suppose the claim is true. Then, since the radius of any slice for $t \geq t_0 = n - \sqrt{D^2 - n}$ is $\Omega(\sqrt{n})$, this is at most $1 - (1 - O(1/n^{1.5}))^n = O(1/\sqrt{n})$ fraction of the slice. Each point in the annulus has a probability of $O(1/\sqrt{n})$ of making a move to the base with no symmetric move away. Thus, effectively, the process can be viewed as a biased random walk on an interval of length $D \geq C_1\sqrt{n}$ with the projection of each step along e_1 is balanced with probability $1 - \Omega\left(\frac{1}{\sqrt{n}}\right)$ and is $O\left(\frac{1}{n}\right)$ biased towards the base with probability $O\left(\frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}}\right)$. Thus the drift is $+O\left(\frac{1}{n^2}\right)$ and it takes $\Omega\left(n^2D\right)$ steps to cover the $\Omega(D)$ distance to get within distance 1 of the base along e_1. This is necessary to have total variation distance less than (say) 0.1 from the target speedy distribution over the body. □

Proof. W now prove the claim. We need to argue that starting at the distribution at time t in each slice is unimodal, i.e., radially monotonic nonincreasing going out from the center. Consider two points x, y on the same slice with x closer to the boundary. WLOG we can assume they are on the same radial line from the center. Also,$\ell(x) \le \ell(y)$. Let $B(x) = K_D \cap (x + \delta B)$. Every point $z \in B(y) \setminus B(x)$ has $\ell(z) \ge \ell(w)$ for any point in $w \in B(x) \setminus B(y)$. Suppose the current distribution is p_t, which is a radially monotonic nonincreasing function. Then,

$$p_{t+1}(y) - p_{t+1}(x) = \int_{z \in B(y)} \frac{p_t(z)}{\mathrm{vol}(\delta B)} dz + (1 - \ell(y))p_t(y) - \int_{z \in B(x)} \frac{p_t(z)}{\mathrm{vol}(\delta B)} dz - (1 - \ell(x))p_t(x)$$

The above expression is minimized by choosing p_t to be as uniform as possible (under the constraint of monotonicity, i.e., $p_t(y) \ge p_t(x)$ and $\forall z \in B(y) \setminus B(x), w \in B(x) \setminus B(y)$, $p_t(z) \ge p_t(w)$) in particular we can set $p_t(z) = p_t$ to be constant in $B(x) \cup B(y)$. Then,

$$\begin{aligned} p_{t+1}(y) - p_{t+1}(x) &\ge p_t \ell(y) + (1 - \ell(y))p_t - p_t \ell(x) - (1 - \ell(x))p_t \\ &= 0. \end{aligned}$$

Thus, the new density maintains monotonicity as claimed. □

Acknowledgement. This work was supported in part by NSF Awards CCF-1563838, CCF-1717349 and CCF-1740551. We thank Ravi Kannan and Ronen Eldan for inspiring discussions, and Ronen for his wonderful invention of stochastic localization.

References

[1] Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 237–245. ACM, 2015.

[2] Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2013.

[3] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM J. Comput.*, 41(6):1704–1721, 2012.

[4] Silouanos Brazitikos, Apostolos Giannopoulos, Petros Valettas, and Beatrice-Helen Vritsiou. *Geometry of isotropic convex bodies*, volume 196. American Mathematical Society Providence, 2014.

[5] Jeff Cheeger. *A lower bound for the smallest eigenvalue of the Laplacian*, pages 195–199. Princeton Univ. Press, 1969.

[6] P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 08 1996.

[7] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23:532–569, 2013.

[8] Alan Frieze and Ravi Kannan. Log-sobolev inequalities and sampling from log-concave distributions. *Ann. Appl. Probab.*, 9(1):14–26, 02 1999.

[9] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105(4):843–854, 1983.

[10] Olivier Guedon and Emanuel Milman. Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. *Geometric and Functional Analysis*, 21(5):1043–1068, 2011.

[11] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13:541–559, 1995.

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

[13] Ravi Kannan, László Lovász, and Ravi Montenegro. Blocking conductance and mixing in random walks. *Combinatorics, Probability & Computing*, 15(4):541–570, 2006.

[14] Michel Ledoux. A simple analytic proof of an inequality by p. buser. *Proceedings of the American mathematical society*, 121(3):951–959, 1994.

[15] Michel Ledoux. Concentration of measure and logarithmic sobolev inequalities. *SÃ©minaire de probabilitÃ©s de Strasbourg*, 33:120–216, 1999.

[16] Yin Tat Lee and He Sun. Constructing linear-sized spectral sparsification in almost-linear time. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 250–269. IEEE, 2015.

[17] Yin Tat Lee and He Sun. An sdp-based algorithm for linear-sized spectral sparsification. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 678–687. ACM, 2017.

[18] Yin Tat Lee and Santosh Srinivas Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *Proc. of IEEE FOCS*, 2017.

[19] L. Lovász and S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures and Algorithms*, 30(3):307–358, 2007.

[20] G. Paouris. Concentration of mass on convex bodies. *Geometric and Functional Analysis*, 16:1021–1049, 2006.

[21] Nikhil Srivastava, Roman Vershynin, et al. Covariance estimation for distributions with 2+eps moments. *The Annals of Probability*, 41(5):3081–3111, 2013.

[22] V.G.Maz'ja. Classes of domains and imbedding theorems for function spaces. *Dokl. Acad. Nauk SSSR (Engl. transl. Soviet Math. Dokl., 1 (1961) 882-885)*, 3:527–530, 1960.