Eldan’s Stochastic Localization and the KLS Hyperplane Conjecture:
An Improved Lower Bound for Expansion

Yin Tat Lee, Santosh S. Vempala

March 31, 2017

Abstract

We show that the KLS constant for n-dimensional isotropic logconcave measures is $O(n^{1/4})$, improving on the current best bound of $O(n^{1/3} \sqrt{\log n})$. As corollaries we obtain the same improved bound on the thin-shell estimate, Poincaré constant and Lipschitz concentration constant and an alternative proof of this bound for the isotropic constant; it also follows that the ball walk for sampling from an isotropic logconcave density in $\mathbb{R}^n$ converges in $O^*(n^{2.5})$ steps from a warm start.
1 Introduction

The isoperimetry of a subset is the ratio of the measure of the boundary of the subset to the measure of the subset or its complement, whichever is smaller. The minimum such ratio over all subsets is the Cheeger constant, also known as expansion or isoperimetric coefficient. This fundamental constant appears in many settings, e.g., graphs and convex bodies and plays an essential role in many lines of study.

In the geometric setting, the KLS hyperplane conjecture [20] asserts that for any distribution with a logconcave density, the minimum expansion is approximated by that of a halfspace, up to a universal constant factor. Thus, if the conjecture is true, the Cheeger constant can be essentially determined simply by examining hyperplane cuts. More precisely, here is the statement. We use $c, C$ for absolute constants, and $\|A\|_2$ for the spectral/operator norm of a matrix $A$.

**Conjecture 1 ([20]).** For any logconcave density $p$ in $\mathbb{R}^n$ with covariance matrix $A$,

$$\frac{1}{\psi_p} \overset{\text{def}}{=} \inf_{S \subseteq \mathbb{R}^n} \frac{\int_{S^c} p(x)dx}{\min \{\int_S p(s)dx, \int_{\mathbb{R}^n \setminus S} p(x)dx\}} \geq \frac{c}{\sqrt{\|A\|_2}}.$$  

For an isotropic logconcave density (all eigenvalues of its covariance matrix are equal to 1), the conjectured isoperimetric ratio is an absolute constant. Note that the isoperimetric constant or KLS constant $\psi_p$ is the reciprocal of the minimum expansion or Cheeger constant (this will be more convenient for comparisons with other constants). The conjecture was formulated by Kannan, Lovász and Simonovits in the course of their study of the convergence of a random process (the ball walk) in a convex body. They proved the following weaker bound.

**Theorem 2 ([20]).** For any logconcave density $p$ in $\mathbb{R}^n$ with covariance matrix $A$, the KLS constant satisfies

$$\psi_p \leq C \sqrt{\text{Tr}(A)}.$$  

For an isotropic distribution, the theorem gives a bound of $O(\sqrt{n})$, while the conjecture says $O(1)$. The conjecture has several important consequences. For example, it implies that the ball walk mixes in $O^*(n^2)$ steps from a warm start in any isotropic convex body (or logconcave density) in $\mathbb{R}^n$; this is the best possible bound, and is tight e.g., for a hypercube. The KLS conjecture has become central to modern asymptotic convex geometry. It is equivalent to a bound on the spectral gap of isotropic logconcave functions [25]. Although it was formulated due to an algorithmic motivation, it implies several well-known conjectures in asymptotic convex geometry. We describe these next.

The thin-shell conjecture (also known as the variance hypothesis) [32, 5] says the following.

**Conjecture 3 (Thin-shell).** For a random point $X$ from an isotropic logconcave density $p$ in $\mathbb{R}^n$,

$$\sigma_p^2 \overset{\text{def}}{=} \mathbb{E}((\|X\| - \sqrt{n})^2) = O(1).$$

It implies that a random point $X$ from an isotropic logconcave density lies in a constant-width annulus (a thin shell) with constant probability. Noting that

$$\sigma_p^2 = \mathbb{E}((\|X\| - \sqrt{n})^2) \leq \frac{1}{n} \text{Var}(\|X\|^2) \leq C \sigma_p^2,$$

the conjecture is equivalent to asserting that $\text{Var}(\|X\|^2) = O(n)$ for an isotropic logconcave density. The following connection is well-known: $\sigma_p \leq C \psi_p$. The current best bound is $\sigma_p \leq n^{\frac{4}{3}}$ by Guedon and Milman [19], improving on a line of work that started with Klartag [23, 24, 17]. Eldan [13] has shown that the reverse inequality holds approximately, in a worst-case sense, namely the worst possible KLS constant over all isotropic logconcave densities in $\mathbb{R}^n$ is bounded by the thin-shell estimate to within roughly a logarithmic factor in the dimension. This yields the current best bound of $\psi_p \leq n^{\frac{4}{3}} \sqrt{\log n}$. A weaker inequality was shown earlier by Bobkov [4] (see also [33]).

The slicing conjecture, also called the hyperplane conjecture [7, 3] is the following.
Conjecture 4 (Slicing/Isotropic constant). Any convex body of unit volume in $\mathbb{R}^n$ contains a hyperplane section of at least constant volume. Equivalently, for any convex body $K$ of unit volume with covariance matrix $L_K^2 I$, the isotropic constant $L_K = O(1)$.

The isotropic constant of a general isotropic logconcave density $p$ with covariance a multiple of the identity is defined as $L_p = p(0)^{1/n}$. The best current bound is $L_p = O(n^{1/4})$, due to Klartag [22], improving on Bourgain’s bound of $L_p = O(n^{1/4} \log n)$ [6]. The study of this conjecture has played an influential role in the development of convex geometry over the past several decades. It was shown by Ball that the KLS conjecture implies the slicing conjecture. More recently, Eldan and Klartag [14] showed that the thin shell conjecture implies slicing, and therefore an alternative (and stronger) proof that KLS implies slicing: $L_p \leq C \sigma_p \leq C' \psi_p$.

Finally, we state a few applications of the KLS bound.

**Theorem 5 (Poincaré constant [36, 10]).** For any isotropic logconcave density $p$ in $\mathbb{R}^n$, we have

$$
\sup_{g \text{ smooth}} \frac{\text{Var}_p(g(x))}{\mathbb{E}_p(\|\nabla g(x)\|^2_2)} = O(\psi_p^2).
$$

**Theorem 6 (Lipschitz concentration [18]).** For any $L$-Lipschitz function $g$ in $\mathbb{R}^n$, and isotropic logconcave density $p$,

$$
\mathbb{P}_{X \sim p}(\|g(x) - \mathbb{E}g\| > \psi_p L t) \leq e^{-\Omega(t)}.
$$

**Theorem 7 (Central Limit Theorem [32]).** Let $K$ be an isotropic symmetric convex set. Let $g_0(s) = \text{vol}(K \cap \{x^T \theta = s\})$ and $g(s) = \frac{1}{\sqrt{2\pi}} \exp(-s^2/2)$. There are universal constants $c_1, c_2 > 0$ such that for any $\delta > 0$, we have

$$
\text{vol}\left(\left\{ \theta \in S^{n-1} : \left| \int_{-\delta}^{\delta} g_0(s)ds - \int_{-\delta}^{\delta} g(s)ds \right| \leq c_1(\delta + \psi_K) \text{ for every } t \in \mathbb{R} \right\}\right) \geq 1 - ne^{-c_2 \delta^2 n}.
$$

For more background on these conjectures, we refer the reader to [9, 1, 2].

### 1.1 Results

We prove the following bound, conjectured in this form in [35].

**Theorem 8.** For any logconcave density $p$ in $\mathbb{R}^n$, with covariance matrix $A$,

$$
\psi_p \leq C (\text{Tr} (A^2))^{1/4}.
$$

For isotropic $p$, this gives a bound of $\psi_p \leq C n^{1/4}$, improving on the current best bound. The following corollary is immediate. We note that it also gives an alternative proof of the central limit theorem for logconcave distributions, via Bobkov’s theorem [4].

**Corollary 9.** For any logconcave density $p$ in $\mathbb{R}^n$, the isotropic (slicing) constant $L_p$ and the thin-shell constant $\sigma_p$ are bounded by $O(n^{1/4})$.

We mention an algorithmic consequence. The ball walk in a convex body $K \subset \mathbb{R}^n$ starts at some point $x_0$ in its interior and at each step picks a uniform random point in the ball of fixed radius $\delta$ centered at the current point, and goes to the point if it lies in $K$. The process converges to the uniform distribution over $K$ in the limit. Understanding the precise rate of convergence is a major open problem with a long line of work and directly motivated the KLS conjecture [26, 27, 28, 20, 21, 29, 30]. Our improvement for the KLS constant gives the following bound on the rate of convergence.

**Corollary 10.** The mixing time of the ball walk to sample from an isotropic logconcave density from a warm start is $O^* (n^{2.5})$. 

2
1.2 Approach

The KLS conjecture is true for Gaussian distributions. More generally, for any distribution whose density function is the product of the Gaussian density for $N(0, \sigma^2 I)$ and any logconcave function, it is known that the expansion is $\Omega(1/\sigma)$ [11]. This fact is used crucially in the Gaussian cooling algorithm of [12] for computing the volume of a convex body by starting with a standard Gaussian restricted to a convex body and gradually making the variance of the Gaussian large enough that it is effectively uniform over the convex body of interest. Our overall strategy is similar in spirit — we start with an arbitrary isotropic logconcave density and gradually introduce a Gaussian term in the density of smaller and smaller variance. The isoperimetry of the resulting distribution after sufficient time will be very good since it has a large Gaussian factor. And crucially, it can be related to the isoperimetry of the initial distribution. To achieve the latter, we would like to maintain the measure of a fixed subset close to its initial value as the distribution changes. For this, our proof uses the localization approach to proving high-dimensional inequalities [28, 20], and in particular, the elegant stochastic version introduced by Eldan [13] and used in subsequent papers [16, 15].

We fix a subset $E$ of the original space with measure one half according to the original logconcave distribution (it suffices to consider such subsets to bound the isoperimetric constant). In standard localization, we then repeatedly bisect space using a hyperplane that preserves the volume fraction of $E$. The limit of this process is a partition into 1-dimensional logconcave measures (“needles”), for which inequalities are much easier to prove. This approach runs into major difficulties for proving the KLS conjecture. While the original measure might be isotropic, the 1-dimensional measures could, in principle, have variances roughly equal to the trace of the original covariance (i.e., long thin needles), for which only much weaker inequalities hold. Stochastic localization can be viewed as the continuous time version of this process, where at each step, we pick a random direction and multiply the current density with a linear function along the chosen direction. Over time, the density can be viewed as a spherical Gaussian times a logconcave function, with the Gaussian gradually reducing in variance. When the Gaussian becomes sufficiently small in variance, then the overall distribution has good isoperimetric coefficient, determined by the inverse of the Gaussian standard deviation (such an inequality can be shown using standard localization, as in [11]). An important property of the infinitesimal change at each step is balance — the density at time $t$ is a martingale and therefore the expected measure of any subset is the same as the original measure. Over time, the measure of a set $E$ is a random quantity that deviates from its original value of $\frac{1}{2}$ over time. The main question then is: what direction to use at each step so that (a) the measure of $E$ remains bounded and (b) the Gaussian part of the density has small variance. We show that the simplest choice, namely a pure random direction chosen from the uniform distribution suffices. The analysis needs a potential function that grows slowly but still maintains good control over the spectral norm of the current covariance matrix. The direct choice of $\|A_t\|_2$ where $A_t$ is the covariance matrix of the distribution at time $t$, is hard to control. We use $\text{Tr}(A_t^2)$. This gives us the improved bound of $O(n^{1/4})$.

2 Preliminaries

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

2.1 Stochastic calculus

In this paper, we only consider stochastic processes given by stochastic differential equations. 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

$$[x]_t = \lim_{|P| \to 0} \sum_{n=1}^{\infty} (x_{\tau_n} - x_{\tau_{n-1}})^2$$

and

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

where $P = \{0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots \uparrow t\}$ 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 via polarization as

$$|x,y|_t = \frac{1}{4} (|x+y|_t - |x-y|_t).$$

For example, 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 $d[x^i,x^j]_t = (\Sigma(x_t)M^T(y_t))_{ij}dt$.

**Lemma 11** (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 12** (Reflection principle). Given a Wiener process $W(t)$ and $a, t \geq 0$, then we have that

$$P(\sup_{0 \leq s \leq t} W(s) \geq a) = 2P(W(t) \geq a).$$

Second, a decomposition lemma for continuous martingales.

**Theorem 13** (Dambis, Dubins-Schwarz theorem). Every continuous local martingale $M_t$ is of the form

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

where $W_s$ is a Wiener process.

### 2.2 Logconcave functions

**Lemma 14** (Dinghas; Prékopa; Leindler). The convolution of two logconcave functions is also logconcave; in particular, any linear transformation or any marginal of a logconcave density is logconcave.

The next lemma about logconcave densities is folklore, see e.g., [31].

**Lemma 15** (Logconcave moments). Given a logconcave density $p$ in $\mathbb{R}^n$, and any positive integer $k$,

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

The following elementary concentration lemma is also well-known (this version is from [31]).

**Lemma 16** (Logconcave concentration). For any isotropic logconcave density $p$ in $\mathbb{R}^n$, and any $t > 0$,

$$P_{x \sim p}(\|x\| > t\sqrt{n}) \leq e^{-t^2/2}.$$  

To prove a lower bound on the expansion, it suffices to consider subsets of measure 1/2. This follows from the concavity of the isoperimetric profile. We quote a theorem from [33, Thm 1.8], which applies even more generally to Riemannian manifolds under suitable convexity-type assumptions.

**Theorem 17.** The Cheeger constant of any logconcave density is achieved by a subset of measure 1/2.
3 Eldan’s stochastic localization

In this section, we consider a variant of the stochastic localization scheme introduced in [13]. In discrete localization, the idea would be to restrict the distribution with a random halfspace and repeat this process. In stochastic localization, this discrete step is replaced by infinitesimal steps, each of which is a renormalization with a linear function in a random direction. One might view this informally as an averaging over infinitesimal needles. The discrete time equivalent would be $p_{t+1}(x) = p_t(x)(1 + \sqrt{h(x - \mu_t)^T w})$ for a sufficiently small $h$ and random Gaussian vector $w$. Using the approximation $1 + y \sim e^{y - \frac{1}{2} y^2}$, we see that over time this process introduces a negative quadratic factor in the exponent, which will be the Gaussian factor. As time tends to $\infty$, the distribution tends to a more and more concentrated Gaussian and eventually a delta function, at which point any subset has measure either 0 or 1. The idea of the proof is to stop at a time that is large enough to have a strong Gaussian factor in the density, but small enough to ensure that the measure of a set is not changed by more than a constant.

3.1 The process and its basic properties

Given a distribution with logconcave density $p(x)$, we start at time $t = 0$ with this distribution and at each time $t > 0$, we apply an infinitesimal change to the density. This is done by picking a random direction from a standard Gaussian.

In order to construct the stochastic process, we assume that the support of $p$ is contained in a ball of radius $R > n$. There is only exponentially small probability outside this ball, at most $e^{-cR}$ by Lemma 16. Moreover, since by Theorem 17, we only need to consider subsets of measure $1/2$, this truncation does not affect the KLS constant of the distribution.

Definition 18. Given a logconcave distribution $p$, we define the following stochastic differential equation:

$$c_0 = 0, \quad dc_t = dW_t + \mu_t dt, \quad (3.1)$$

where the probability distribution $p_t$, the mean $\mu_t$ and the covariance $A_t$ are defined by

$$p_t(x) = \frac{e^{c^T x - \frac{1}{2} ||x||^2} p(x)}{\int_{\mathbb{R}^n} e^{c^T y - \frac{1}{2} ||y||^2} p(y) dy}, \quad \mu_t = \mathbb{E}_{x \sim p_t} x, \quad A_t = \mathbb{E}_{x \sim p_t} (x - \mu_t)(x - \mu_t)^T.$$

Since $\mu_t$ is a bounded function that is Lipschitz with respect to $c$ and $t$, standard theorems (e.g. [34, Sec 5.2]) show the existence and uniqueness of the solution in time $[0, T]$ for any $T > 0$. We defer all proofs for statements in this section, considered standard in stochastic calculus, to Section 5. Now we proceed to analyzing the process and how its parameters evolve. Roughly speaking, the first lemma below says that the stochastic process is the same as continuously multiplying $p_t(x)$ by a random infinitesimally small linear function.

Lemma 19. We have that $dp_t(x) = (x - \mu_t)^T dW_t p_t(x)$ for any $x \in \mathbb{R}^n$.

By considering the derivative $d\log p_t(x)$, we see that applying $dp_t(x)$ as in the lemma above results in the distribution $p_t(x)$, with the Gaussian term in the density:

$$d \log p_t(x) = \frac{dp_t(x)}{p_t(x)} - \frac{1}{2} \frac{d|p_t(x)|}{p_t(x)^2} = (x - \mu_t)^T dW_t - \frac{1}{2} (x - \mu_t)^T (x - \mu_t) dt$$

$$= x^T dc_t - \frac{1}{2} ||x||^2 dt + g(t)$$

where the last term is independent of $x$ and the first two terms explain the form of $p_t(x)$ and the appearance of the Gaussian.

Next we analyze the change of the covariance matrix.

Lemma 20. We have that $dA_t = \int_{\mathbb{R}^n} (x - \mu_t)^T A_t (x - \mu_t)^T dW_t p_t(x) dx - A_t^2 dt.$
3.2 Bounding expansion

Our goal is to bound the expansion by the spectral norm of the covariance matrix at time $t$. First, we bound the measure of a set of initial measure $\frac{1}{2}$.

**Lemma 21.** For any set $E \subset \mathbb{R}^n$ with $\int_E p(x)dx = \frac{1}{2}$ and $t \geq 0$, we have that

$$
\mathbb{P}\left(\frac{1}{4} \leq \int_E p_t(x)dx \leq \frac{3}{4}\right) \geq \frac{9}{10} - \mathbb{P}\left(\int_{0}^{t} \|A_s\|_2 ds \geq \frac{1}{64}\right).
$$

**Proof.** Let $g_t = \int_E p_t(x)dx$. Then, we have that

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

where the integral might not be 0 because it is over the subset $E$ and not all of $\mathbb{R}^n$. Hence, we have,

$$
d|g|_t = \left\| \int_E (x - \mu_t)p_t(x)dx \right\|_2 dt = \max_{\|\zeta\|_2 \leq 1} \left( \int_E \zeta^T (x - \mu_t)p_t(x)dx \int_E p_t(x)dx dt \right)
$$

$$
= \max_{\|\zeta\|_2 \leq 1} (\zeta^T A_t \zeta) dt = \|A_t\|_2 dt.
$$

Hence, we have that $\frac{dg_t}{dt} \leq \|A_t\|_2$. By the Dambis, Dubins-Schwarz theorem, there exists a Wiener process $W_t$ such that $g_t - g_0$ has the same distribution as $W_t$. Using $g_0 = \frac{1}{2}$, we have that

$$
\mathbb{P}\left(\frac{1}{4} \leq g_t \leq \frac{3}{4}\right) = \mathbb{P}\left(\frac{1}{4} \leq W_t \leq \frac{3}{4}\right) \geq 1 - \mathbb{P}(\max_{0 \leq s \leq t} |W_s| > \frac{1}{4}) - \mathbb{P}(|g|_t > \frac{1}{64})
$$

$$
\overset{(1)}{\geq} 1 - 4\mathbb{P}(W_t > \frac{1}{4}) - \mathbb{P}(|g|_t > \frac{1}{64})
$$

$$
\overset{(2)}{\geq} \frac{9}{10} - \mathbb{P}(|g|_t > \frac{1}{64})
$$

where we used reflection principle for 1-dimensional Brownian motion in (1) and the concentration of normal distribution in (2), namely $\mathbb{P}_{x \sim N(0,1)}(x > 2) \leq 0.0228$.

**Theorem 22** (Brascamp-Lieb [8]). Let $\gamma : \mathbb{R}^n \to \mathbb{R}_+$ be the standard Gaussian density in $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be any logconcave function. Define the density function $h$ as follows:

$$
h(x) = \frac{f(x)\gamma(x)}{\int_{\mathbb{R}^n} f(y)\gamma(y) dy}.
$$

Fix a unit vector $v \in \mathbb{R}^n$, let $\mu = \mathbb{E}_h(x)$. Then, for any $\alpha \geq 1$, $\mathbb{E}_h(|v^T(x - \mu)|^\alpha) \leq \mathbb{E}_\gamma(|v^T x|^\alpha)$.

Using the above, the following isoperimetric inequality was proved in [11] and was also used in [13].

**Theorem 23** ([11, Thm. 4.4]). Let $h(x) = f(x)e^{-\frac{1}{2}x^TBx}/\int f(y)e^{-\frac{1}{2}y^TBy}dy$ where $f : \mathbb{R}^n \to \mathbb{R}_+$ is an integrable logconcave function and $B$ is positive definite. Then $h$ is logconcave and for any measurable subset $S$ of $\mathbb{R}^n$,

$$
\int_{\partial S} h(x)dx = \Omega \left(\|B^{-1}\|_2^{-\frac{1}{2}}\right) \min \left\{ \int_S h(x)dx, \int_{\mathbb{R}^n \setminus S} h(x)dx \right\}.
$$

In other words, the expansion of $h$ is $\Omega \left(\|B^{-1}\|_2^{-\frac{1}{2}}\right)$. 

6
Proof. The proof uses the localization lemma to reduce the statement to a 1-dimensional statement about a Gaussian times a logconcave density, where the Gaussian is a projection of the Gaussian \(N(0, B^{-1})\) (but the logconcave function might be different as the limit of localization is the original function along an interval times an exponential function). We then apply the Brascamp-Lieb inequality in one dimension (Theorem 22) to prove that for the resulting one-dimensional distribution, the variance is at most that of the Gaussian, therefore at most \(\|B^{-1}\|\). The isoperimetric constant is bounded by the inverse of the standard deviation times a constant. The complete proof, in more general terms, is carried out in [11, Thm. 4.4]. \(\square\)

We can now prove a bound on the expansion.

**Lemma 24.** Given a logconcave distribution \(p\). Let \(A_t\) be defined by Definition 18 using initial distribution \(p\). Suppose that there is \(T > 0\) such that

\[
P \left( \int_0^T \|A_s\|_2^2 \, ds \leq \frac{1}{64} \right) \geq \frac{3}{4}
\]

Then, we have that \(\psi_p = \Omega(\sqrt{T})\).

**Proof.** By Milman’s theorem [33], it suffices to consider subsets of measure \(\frac{1}{2}\). Consider any measurable subset \(E \subset \mathbb{R}^n\) of initial measure \(\frac{1}{2}\). By Lemma 19, \(p_t\) is a martingale and therefore

\[
\int_{\partial E} p(x) \, dx = \int_{\partial E} p_0(x) \, dx = \mathbb{E} \left( \int_{\partial E} p_t(x) \, dx \right).
\]

Next, by the definition of \(p_T\) (3.1), we have that \(p_T(x) \propto e^{\frac{1}{2}x^T B^{-1} x} \|x\|^2 p(x)\) and Theorem 23 shows that the expansion of \(E\) is \(\Omega(\sqrt{T})\). Hence, we have

\[
\int_{\partial E} p(x) \, dx = \mathbb{E} \left( \int_E p_T(x) \, dx \right) \geq \Omega(\sqrt{T}) \mathbb{E} \left( \min \left( \int_E p_T(x) \, dx, \int_{E^c} p_T(x) \, dx \right) \right) \geq \Omega(\sqrt{T}) \mathbb{E} \left( \int_E p_T(x) \, dx \leq \frac{3}{4} \right) \geq 21 \Omega(\sqrt{T}) \left( \frac{9}{10} - \mathbb{P} \left( \int_0^T \|A_s\|_2^2 \, ds \geq \frac{1}{64} \right) \right) = \Omega(\sqrt{T})
\]

where we used the assumption at the end. Using Theorem 17, this shows that \(\psi_p = \Omega(\sqrt{T})\). \(\square\)

## 4 Controlling \(A_t\) via the potential \(\text{Tr}(A_t^2)\)

### 4.1 Third moment bounds

Here are two key lemmas about the third-order tensor of a log-concave distribution.

**Lemma 25.** Given a logconcave distribution \(p\) with mean \(\mu\) and covariance \(A\). For any positive semi-definite matrix \(C\), we have that

\[
\|E_{x \sim p} (x - \mu)(x - \mu)^T C(x - \mu)\|_2 = O \left( \|A\|_2^{1/2} \text{Tr} \left( A^{1/2} C A^{1/2} \right) \right).
\]

**Proof.** We first prove the case \(C = v_3^T\). Taking \(y = A^{-1/2}(x - \mu)\) and \(w = A^{1/2}v\). Then, \(y\) follows an isotropic logconcave distribution \(\tilde{p}\) and hence

\[
\left\| E_{y \sim \tilde{p}} \left( y^T w \right)^2 \right\|_2 = \left\| E_{y \sim \tilde{p}} A^{1/2} y^T A^{1/2} w \right\|_2 = \max_{\|\zeta\|_2 \leq 1} E_{y \sim \tilde{p}} \left( A^{1/2} y^T \zeta \right) (y^T w)^2 \leq \max_{\|\zeta\|_2 \leq 1} \sqrt{E_{y \sim \tilde{p}} \left( (A^{1/2} y^T \zeta) \right)^2} \sqrt{E_{y \sim \tilde{p}} (y^T w)^4} = O \left( \|A\|_2^{1/2} \|w\|_2^2 \right)
\]

7
where we used the fact that for a fixed \( w \), \( y^T w \) has a one-dimensional logconcave distribution (Lemma 14) and hence Lemma 15 shows that
\[
\mathbb{E}_{y \sim \rho} (y^T w)^4 = O(1) \left( \mathbb{E}_{y \sim \rho} (y^T w)^2 \right)^2 = O(\|w\|_2^4).
\]

For a general PSD matrix \( C \), we write \( C = \sum \lambda_i v_i v_i^T \) where \( \lambda_i \geq 0 \), \( v_i \) are eigenvalues and eigenvectors of \( C \). Hence, we have that
\[
\| \mathbb{E}_{x \sim p}(x - \mu) (x - \mu)^T C(x - \mu) \|_2 \leq \sum \lambda_i \| \mathbb{E}_{x \sim p}(x - \mu) (x - \mu)^T v_i v_i^T (x - \mu) \|_2 \leq O(1) \sum \lambda_i \| A \|_2^{1/2} \| A_{1/2} v_i \|_2^2
\]
\[
= O(1) \| A \|_2^{1/2} \sum \text{Tr} \left( A_{1/2} \lambda_i v_i v_i^T A_{1/2}^2 \right) = O(1) \| A \|_2^{1/2} \text{Tr} \left( A_{1/2} C A_{1/2}^2 \right).
\]

**Lemma 26.** Given a logconcave distribution \( p \) with mean \( \mu \) and covariance \( A \). We have
\[
\mathbb{E}_{x,y \sim p} |\langle x - \mu, y - \mu \rangle|^3 = O \left( \text{Tr} \left( A^2 \right)^{3/2} \right).
\]

**Proof.** Without loss of generality, we assume \( \mu = 0 \). For a fixed \( x \) and random \( y \), \( \langle x, y \rangle \) follows a one-dimensional logconcave distribution (Lemma 14) and hence Lemma 15 shows that
\[
\mathbb{E}_{y \sim p} |\langle x, y \rangle|^3 \leq O(1) \left( \mathbb{E}_{y \sim p} (x, y)^2 \right)^{3/2} = O \left( x^T A x \right)^{3/2}.
\]

Next, we note that \( A_{1/2} x \) follows a logconcave distribution (Lemma 14) and hence Lemma 15 shows that
\[
\mathbb{E}_{x,y \sim p} |\langle x, y \rangle|^3 = O(1) \mathbb{E}_{x \sim p} \| A_{1/2} x \|^3 \leq O(1) \left( \mathbb{E}_{x \sim p} \| A_{1/2} x \|^2 \right)^{3/2} = O \left( \text{Tr} \left( A^2 \right)^{3/2} \right).
\]

**4.2 Analysis of \( A_t \)**

Using Itô’s formula and Lemma 20, we compute the derivatives of \( \text{Tr} A_t^2 \).

**Lemma 27.** Let \( A_t \) be defined by Definition 18. We have that
\[
d\text{Tr} A_t^2 = 2 \mathbb{E}_{x \sim p_t} (x - \mu_t)^T A_t (x - \mu_t)(x - \mu_t)^T dW_t - 2 \text{Tr}(A_t^3) dt + \mathbb{E}_{x,y \sim p_t} (x - \mu_t)^T(y - \mu_t)^3 dt.
\]

**Lemma 28.** Given a logconcave distribution \( p \) with covariance matrix \( A \) s.t. \( \text{Tr} A^2 = n \). Let \( A_t \) defined by Definition 18 using initial distribution \( p \). There is a universal constant \( c_1 \) such that
\[
P \left( \max_{t \in [0, T]} \text{Tr} \left( A_t^2 \right) \geq 8 n \right) \leq 0.01 \quad \text{with} \quad T = \frac{c_1}{\sqrt{n}}.
\]

**Proof.** Let \( \Phi_t = \text{Tr} A_t^2 \). By Lemma 27, we have that
\[
d\Phi_t = -2 \text{Tr}(A_t^3) dt + \mathbb{E}_{x,y \sim p_t} ((x - \mu_t)^T(y - \mu_t))^3 dt + 2 \mathbb{E}_{x \sim p_t} (x - \mu_t)^T A_t (x - \mu_t)(x - \mu_t)^T dW_t
\]
\[
\overset{\text{def}}{=} \delta_t dt + v_i^T dW_i.
\]

(4.1)

For the drift term \( \delta_t dt \), Lemma 26 shows that
\[
\delta_t \leq \mathbb{E}_{x,y \sim p_t} ((x - \mu_t)^T(y - \mu_t))^3 = O \left( \text{Tr} \left( A_t^2 \right)^{3/2} \right) \leq C' \Phi_t^{3/2}
\]

(4.2)

for some universal constant \( C' \). Note that we dropped the term \(-2 \text{Tr}(A_t^3)\) since \( A_t \) is positive semidefinite and therefore the term is negative.
For the martingale term $v_t^T dW_t$, we note that

$$
\|v_t\|_2 = \|E_{x \sim p_t} (x - \mu_t) y \|_2 \leq \|A_t\|^{1/2} T \|A_t^2\| \leq \Phi_0^{5/2}.
$$

So the drift term grows roughly as $\Phi^{3/2} t$ while the stochastic term grows as $\Phi^{5/4} \sqrt{t}$. Thus, both bounds (on the drift term and the stochastic term) suggest that for $t$ up to $O \left( \frac{1}{\sqrt{n}} \right)$, the potential $\Phi_t$ remains $O(n)$. We now formalize this, by decoupling the two terms.

Let $f(a) = -\frac{1}{\sqrt{n+2}}$. By (4.1) and Itô’s formula, we have that

$$
df(\Phi_t) = f'(\Phi_t) d\Phi_t + \frac{1}{2} f''(\Phi_t) d[\Phi]_t = \left( \frac{\delta_t}{2 (\Phi_t + n)^{3/2}} - \frac{3}{8} \frac{\|v_t\|_2^2}{(\Phi_t + n)^{3/2}} \right) dt + \frac{1}{2} \frac{v_t^T dW_t}{(\Phi_t + n)^{3/2}}.
$$

where $dY_t = \frac{v_t^T dW_t}{(\Phi_t + n)^{3/2}}$, $Y_t = 0$ and $C'$ is the universal constant in (4.2).

Note that

$$
d[Y]_t = \frac{1}{4} \frac{\|v_t\|_2^2}{(\Phi_t + n)^3} = O(1) \frac{\Phi^{5/2}}{(\Phi_t + n)^3} \leq \frac{C}{\sqrt{n}}.
$$

By Theorem 13, there exists a Wiener process $\tilde{W}_t$ such that $Y_t$ has the same distribution as $\tilde{W}[\gamma]$. Using the reflection principle for 1-dimensional Brownian motion, we have that

$$
P(\max_{t \in [0,T]} Y_t \geq \gamma) \leq P(\max_{t \in [0,T]} \tilde{W}_t \geq \gamma) = 2P(\tilde{W}_t \geq \gamma) \leq 2 \exp\left(-\frac{\gamma^2}{2CT} \right).
$$

Since $\Phi_0 = \|A_p\|^2 = n$, we have that $f(\Phi_0) = -\frac{1}{2n}$ and therefore (4.3) shows that

$$
P(\max_{t \in [0,T]} f(\Phi_t) \geq -\frac{1}{2n} + C'T + \gamma) \leq 2 \exp\left(-\frac{\gamma^2}{2CT} \right).
$$

Putting $T = \frac{1}{256(C' + C)^2} \frac{1}{\sqrt{n}}$ and $\gamma = \frac{1}{4\sqrt{n}}$, we have that

$$
P(\max_{t \in [0,T]} f(\Phi_t) \geq -\frac{1}{3\sqrt{n}}) \leq 2 \exp(-8).
$$

Note that $f(\Phi_t) \geq -\frac{1}{3\sqrt{n}}$ implies that $\Phi_t \geq 8n$. Hence, we have that $P(\max_{t \in [0,T]} \Phi_t \geq 8n) \leq 0.01$. \qed

4.3 Proof of Theorem 8

Proof of Theorem 8. By rescaling, we can assume $\text{Tr} A^2 = n$. By Lemma 28, we have that

$$
P(\max_{s \in [0,t]} \text{Tr} (A_s^2) \leq 8n) \geq 0.99 \text{ with } t = \frac{c_1}{\sqrt{n}}.
$$

Since $\text{Tr} (A_t^2) \leq 8n$ implies that $\|A_t\|_2 \leq \sqrt{8n}$, we have that $P(\int_0^T \|A_s\| ds \leq \frac{1}{8\sqrt{t}}) \geq 0.99$ where $T = \min \left\{ \frac{1}{64\sqrt{8}}, c_1 \right\} \frac{1}{\sqrt{n}}$. Now the theorem follows from Lemma 24. \qed

5 Localization proofs

We begin with the proof of the infinitesimal change in the density.
Proof of Lemma (19). Let \( q_t(x) = e^{c_t^T x - \frac{1}{2} \|x\|^2} p(x) \). By Itô’s formula, applied to \( f(a, t) \equiv e^{a - \frac{1}{2} \|x\|^2} p(x) \) with \( a = c_t^T x \), we have that

\[
dq_t(x) = \frac{df(a, t)}{da} dc_t^T x + \frac{df(a, t)}{dt} dt + \frac{1}{2} d^2 f(a, t) da^2 d[c_t^T x]_t + \frac{1}{2} d^2 f(a, t) dt^2 d[t]_t + \frac{1}{2} \cdot 2 \cdot \frac{d^2 f(a, t)}{dadt} d[c_t^T x, t]_t
\]

\[
= \left( dc_t^T x - \frac{1}{2} \|x\|^2 dt + \frac{1}{2} d[c_t^T x]_t \right) q_t(x).
\]

By the definition of \( c_t \), we have \( dc_t^T x = \langle dW_t + \mu_t dt, x \rangle \). The quadratic variation of \( c_t^T x \) is \( d[c_t^T x]_t = \langle x, x \rangle dt \). The other two quadratic variation terms are zero. Therefore, this gives

\[
dq_t(x) = \langle dW_t + \mu_t dt, x \rangle q_t(x).
\]

Let \( V_t = \int_{\mathbb{R}^n} q_t(y)dy \). Then, we have

\[
dV_t = \int_{\mathbb{R}^n} dq_t(y)dy = \int_{\mathbb{R}^n} \langle dW_t + \mu_t dt, y \rangle q_t(y)dy = V_t \langle dW_t + \mu_t dt, \mu_t \rangle.
\]

By Itô’s formula, we have that

\[
dV_t^{-1} = -\frac{1}{V_t^2} dV_t + \frac{1}{V_t^2} d[V]_t = -V_t^{-1} \langle dW_t + \mu_t dt, \mu_t \rangle + V_t^{-1} \langle \mu_t, \mu_t \rangle dt = -V_t^{-1} \langle dW_t, \mu_t \rangle.
\]

Combining (5.1) and (5.2), we have that

\[
dp_t(x) = d(V_t^{-1} q_t(x)) = q_t(x)dV_t^{-1} + V_t^{-1} dq_t(x) + d[V_t^{-1}, q_t(x)]_t = p_t(x) \langle dW_t, x - \mu_t \rangle.
\]
Using the formula for $W$ given by Lemma 19, we have that
\[ \int_{\mathbb{R}^n} (x - \mu_t)d[\mu_T]_t dt = \int_{\mathbb{R}^n} (x - \mu_t)(x - \mu_t)^T A_t p_t(x) dt dx = A_t^2 dt. \]
Similarly, we have the fifth term $\int_{\mathbb{R}^n} d[\mu_t, p_t(x)]_t (x - \mu_t)^T A_t p_t(x) dt dx = A_t^2 dt$. Combining all the terms, we have that
\[ dA_t = \int_{\mathbb{R}^n} (x - \mu_t)(x - \mu_t)^T dp_t(x) dx - A_t^2 dt. \]

Next is the proof of stochastic derivative of the potential $\Phi_t = \text{Tr}(A_t^2)$.

**Proof of Lemma 27.** Let $\Phi(X) = \text{Tr}(X^2)$. Then the first and second-order directional derivatives of $\Phi$ at $X$ is given by $\frac{\partial \Phi}{\partial X} |_H = 2\text{Tr}(XH)$ and $\frac{\partial^2 \Phi}{\partial X^2} |_{H_1,H_2} = 2\text{Tr}(H_1H_2)$. Using these and Itô’s formula, we have that
\[ d\text{Tr}(A_t^2) = 2\text{Tr}(A_t dA_t) + \sum_{ij} d[A_{ij}, A_{ji}]_t \]
where $A_{ij}$ is the real-valued stochastic process defined by the $(i,j)^{th}$ entry of $A_t$. Using Lemma 20 and Lemma 19, we have that
\[ dA_t = \sum_z \mathbb{E}_{x \sim p_t}(x - \mu_t)(x - \mu_t)^T (x - \mu_t)_z dW_{t,z} - A_t^2 dt \]
(5.3)
where $W_{t,z}$ is the $z^{th}$ coordinate of $W_t$. Therefore,
\[ d[A_{ij}, A_{ji}]_t = \sum_z \left( \mathbb{E}_{x \sim p_t}(x - \mu_t)i(x - \mu_t)_z (x - \mu_t)^T e_z \right) \left( \mathbb{E}_{x \sim p_t}(x - \mu_t)_j (x - \mu_t)_i (x - \mu_t)^T e_z \right) dt \]
\[ = \mathbb{E}_{x,y \sim p_t}(x - \mu_t)i(x - \mu_t)_j (y - \mu_t)_j (y - \mu_t)_i (x - \mu_t)^T (y - \mu_t) dt. \]  
(5.4)
Using the formula for $dA_t$ (5.3) and $d[A_{ij}, A_{ji}]_t$ (5.4), we have that
\[ d\text{Tr}(A_t^2) = 2\mathbb{E}_{x \sim p_t}(x - \mu_t)^T A_t(x - \mu_t)(x - \mu_t)^T dW_t - 2\text{Tr}(A_t^3) dt \]
\[ + \sum_{ij} \mathbb{E}_{x,y \sim p_t}(x - \mu_t)i(x - \mu_t)_j (y - \mu_t)_j (x - \mu_t)_i (x - \mu_t)^T (y - \mu_t) dt \]
\[ = 2\mathbb{E}_{x \sim p_t}(x - \mu_t)^T A_t(x - \mu_t)(x - \mu_t)^T dW_t - 2\text{Tr}(A_t^3) dt + \mathbb{E}_{x,y \sim p_t}((x - \mu_t)^T (y - \mu_t))^3 dt. \]

**Acknowledgement.** We thank Ravi Kannan, Laci Lovász, Assaf Naor and Nisheeth Vishnoi for their continuous support and encouragement. We also thank Sébastien Bubeck, Ben Cousins, Ronen Eldan, Bo’az Klartag, Anup B. Rao and Ting-Kam Leonard Wong for helpful discussions.

**References**


