# An Almost Constant Lower Bound of the Isoperimetric Coefficient in the KLS Conjecture

Geom. Funct. Anal. Vol. 31 (2021) 34–61
https://doi.org/10.1007/s00039-021-00558-4
Published online March 24, 2021

GAFA Geometric And Functional Analysis

# AN ALMOST CONSTANT LOWER BOUND OF THE ISOPERIMETRIC COEFFICIENT IN THE KLS CONJECTURE

Yuansi Chen



**Abstract.** We prove an almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. The lower bound has the dimension dependency $d^{-o_d(1)}$. When the dimension is large enough, our lower bound is tighter than the previous best bound which has the dimension dependency $d^{-1/4}$. Improving the current best lower bound of the isoperimetric coefficient in the KLS conjecture has many implications, including improvements of the current best bounds in Bourgain's slicing conjecture and in the thin-shell conjecture, better concentration inequalities for Lipschitz functions of log-concave measures and better mixing time bounds for MCMC sampling algorithms on log-concave measures.

## 1 Introduction

Given a distribution, the isoperimetric coefficient of a subset is the ratio of the measure of the subset boundary to the minimum of the measures of the subset and its complement. Taking the minimum of such ratios over all subsets defines the isoperimetric coefficient of the distribution, also called the Cheeger isoperimetric coefficient of the distribution.

Kannan, Lovász and Simonovits (KLS) [12] conjecture that for any distribution that is log-concave, the Cheeger isoperimetric coefficient equals to that achieved by half-spaces up to a universal constant factor. If the conjecture is true, the Cheeger isoperimetric coefficient can be determined by going through all the half-spaces instead of all subsets. For this reason, the KLS conjecture is also called the KLS hyperplane conjecture. To make it precise, we start by formally defining log-concave distributions and then we state the conjecture.

A probability density function $p : \mathbb{R}^d \to \mathbb{R}$ is *log-concave* if its logarithm is concave, i.e., for any $x, y \in \mathbb{R}^d \times \mathbb{R}^d$ and for any $\lambda \in [0, 1]$,

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

Common probability distributions such as Gaussian, exponential and logistic are log-concave. This definition also includes any uniform distribution over a *convex set* defined as follows. A subset $K \subset \mathbb{R}^d$ is *convex* if $\forall x, y \in K \times K, z \in [x, y] \implies z \in K$.

The *isoperimetric coefficient* $\psi(p)$ of a density $p$ in $\mathbb{R}^d$ is defined as

$$\psi(p) := \inf_{S \subset \mathbb{R}^d} \frac{p^+(\partial S)}{\min(p(S), p(S^c))} \tag{2}$$

where $p(S) = \int_{x \in S} p(x)dx$ and the boundary measure of the subset is

$$p^+(\partial S) := \liminf_{\epsilon \to 0^+} \frac{p(\{x : \mathbf{d}(x, S) \leq \epsilon\}) - p(S)}{\epsilon},$$

where $\mathbf{d}(x, S)$ is the Euclidean distance between $x$ and the subset $S$.

The KLS conjecture is stated by Kannan, Lovász and Simonovits [12] as follows.

Conjecture 1. *There exists a universal constant $c$, such that for any log-concave density $p$ in $\mathbb{R}^d$, we have*

$$\psi(p) \geq \frac{c}{\sqrt{\rho(p)}},$$

*where $\rho(p)$ is the spectral norm of the covariance matrix of $p$. In other words, $\rho(p) = \|A\|_2$, where $A = \mathrm{Cov}_{X \sim p}(X)$ is the covariance matrix.*

An upper bound of $\psi(p)$ of the same form is relatively easy and it was shown to be achieved by half-spaces [12]. Proving the lower bound on $\psi(p)$ up to some small factors in Conjecture 1 is the main goal of this paper. We say a log-concave density is *isotropic* if its mean $\mathbb{E}_{X \sim p}[X]$ equals to 0 and its covariance $\mathrm{Cov}_{X \sim p}(X)$ equals to $\mathbb{I}_d$. In the case of isotropic log-concave densities, the KLS conjecture states that any isotropic log-concave density has its isoperimetric coefficient lower bounded by a universal constant.

There are many attempts trying to lower bound the Cheeger isoperimetric coefficient in the KLS conjecture. We refer readers to the survey paper by Lee and Vempala [18] for a detailed exposition of these attempts. In particular, the original KLS paper [12] (Theorem 5.1) shows that for any log-concave density $p$ with covariance matrix $A$,

$$\psi(p) \geq \frac{\log(2)}{\sqrt{\mathrm{Tr}(A)}}. \tag{3}$$

The original KLS paper [12] only deals with uniform distributions over convex sets, but their proof techniques can be easily extended to show that the same results hold for all log-concave densities. Remark that Equation (3) implies $\psi(p) \geq \frac{\log(2)}{d^{1/2} \cdot \sqrt{\rho(p)}}$. The current best bound is shown in Lee and Vempala [17], where they show that there exists a universal constant $c$ such that for any log-concave density $p$ with covariance matrix $A$,

$$\psi(p) \geq \frac{c}{\left(\mathrm{Tr}\left(A^2\right)\right)^{1/4}}. \tag{4}$$

It implies that $\psi(p) \geq \frac{c}{d^{1/4} \cdot \sqrt{\rho(p)}}$. Note that in Lee and Vempala [17], their notation of $\psi(p)$ is the reciprocal of ours and it is later switched in Theorem 32 of the survey paper [18] by the same authors. As a result, the above bound is not a misstatement of the results in Lee and Vempala [17] and it is simply translated into our notations. In this paper, we improve the dimension dependency $d^{-1/4}$ to $d^{-o_d(1)}$ in the lower bound of the isoperimetric coefficient.

There are many implications of improving the lower bound in the KLS conjecture. The two closely related conjectures are Bourgain's slicing conjecture [3, 4] and the thin-shell conjecture [2]. It is worth noting that Bourgain [4] stated the slicing conjecture earlier than the introduction of the KLS conjecture. In terms of their connections to the KLS conjecture, Eldan and Klartag [9] proved that the thin-shell conjecture implies Bourgain's slicing conjecture up to a universal constant factor. Later, Eldan [8] showed that the inverse of an lower bound of the isoperimetric coefficient is equivalent to an upper bound of the thin-shell constant in the thin-shell conjecture. Combining these two results, we have that an lower bound in the KLS conjecture implies upper bounds in the thin-shell conjecture and in Bourgain's slicing conjecture.

The current best upper bound of the thin-shell constant has the dimension dependency $d^{1/4}$ due to Lee and Vempala's [17] improvement in the KLS conjecture. The current best bound of the slicing constant in Bourgain's slicing conjecture also has the dimension dependency $d^{1/4}$, proved by Klartag [13] without using the KLS conjecture. Klartag's slicing constant bound is a slight improvement over Bourgain's earlier slicing bound [4] which has the dimension dependency $d^{1/4}\log(d)$. Given the current best bounds in these three conjectures and the relation among them, we conclude that improving the current best lower bound in the KLS conjecture improves the current best bounds for the other two conjectures, as noted in Lee and Vempala [18]. For a detailed exposition of the three conjectures and related results since the introduction of Bourgain's slicing conjecture, we refer readers to Klartag and Milman [14].

Additionally, improving the lower bound in the KLS conjecture also improves concentration inequalities for Lipschitz functions of log-concave measures. It also leads to faster mixing time bounds of Markov chain Monte Carlo (MCMC) sampling algorithms on log-concave measures. Despite the great importance of these results, deriving these results from our new bound in the KLS conjecture is not the main focus of our paper. We refer readers to Milman [20] and Lee and Vempala [18] for more details about the abundant implications of the KLS conjecture.

*Notation* For two sequences $a_n$ and $b_n$ indexed by an integer $n$, we say that $a_n = o_n(b_n)$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$. The Euclidean norm of a vector $x \in \mathbb{R}^d$ is denoted by $\|x\|_2$. The spectral norm of a square matrix $A \in \mathbb{R}^{d\times d}$ is denoted by $\|A\|_2$. The Euclidean ball with center $x$ and radius $r$ is denoted by $\mathbb{B}(x, r)$. For a real number $x \in \mathbb{R}$, we denote its ceiling by $\lceil x \rceil = \min\{m \in \mathbb{Z} \mid m \geq x\}$. We say a density $p$ is more log-concave than a Gaussian density $\varphi$ if $p$ can be written as a product form

$p = \nu \cdot \varphi$ where $\varphi$ is the Gaussian density and $\nu$ is a log-concave function (that is, $\nu$ is proportional to a log-concave density). For a martingale $(M_t,\ t \in \mathbb{R}_+)$, we use $[M]_t$ to denote its *quadratic variation*, defined as

$$[M]_t = \sup_{k \in \mathbb{N}} \sup_{0 \le t_1 \le \cdots \le t_k \le t} \sum_{i=1}^{k} \left( M_{t_i} - M_{t_{i-1}} \right)^2 .$$

## 2 Main results

We prove the following lower bound on the isoperimetric coefficient of any log-concave density.

**Theorem 1.** *There exists a universal constant $c$ such that for any log-concave density $p$ in $\mathbb{R}^d$ and any integer $\ell \ge 1$, we have*

$$\psi(p) \ge \frac{1}{\left[c \cdot \ell \left(\log(d) + 1\right)\right]^{\ell/2} d^{16/\ell} \cdot \sqrt{\rho\left(p\right)}} \tag{5}$$

*where $\rho\left(p\right)$ is the spectral norm of the covariance matrix of $p$.*

As a corollary, take $\ell = \left\lceil \left( \frac{\log(d)}{\log\log(d)} \right)^{1/2} \right\rceil$, then there exists a constant $c'$ such that

$$\psi(p) \ge \frac{1}{d^{c' \left( \frac{\log\log(d)}{\log d} \right)^{1/2}} \cdot \sqrt{\rho\left(p\right)}}.$$

Since $\lim_{d \to \infty} \frac{\log\log(d)}{\log(d)} = 0$, for $d$ large enough, the above lower bound is better than any lower bound of the form $\frac{1}{d^{c''} \sqrt{\rho(p)}}$ ($c''$ is a positive constant) in terms of dimension $d$ dependency.

The proof of the main theorem uses the stochastic localization scheme introduced by Eldan [8]. Eldan uses this stochastic localization scheme to show that the thin shell conjecture is equivalent to the KLS conjecture up to a logarithmic factor. The construction of stochastic localization scheme uses elementary properties of semimartingales and stochastic integration. The main idea of Eldan's proof to derive the KLS conjecture from the thin shell conjecture is to smoothly multiply a Gaussian part to the log-concave density, so that the modified density is more log-concave than a Gaussian density. When the Gaussian part is large enough, one can then easily prove the isoperimetric inequality.

The same scheme was refined in Lee and Vempala [17] to obtain the current best lower bound in the KLS conjecture. Lee and Vempala directly attack the KLS conjecture while following the same stochastic localization scheme to smoothly multiply a Gaussian part to the log-concave density. Their use of a new potential function leads to the current best lower bound in the KLS conjecture. The proof in this paper builds on Lee and Vempala [17]'s refinements of Eldan's method, while it improves

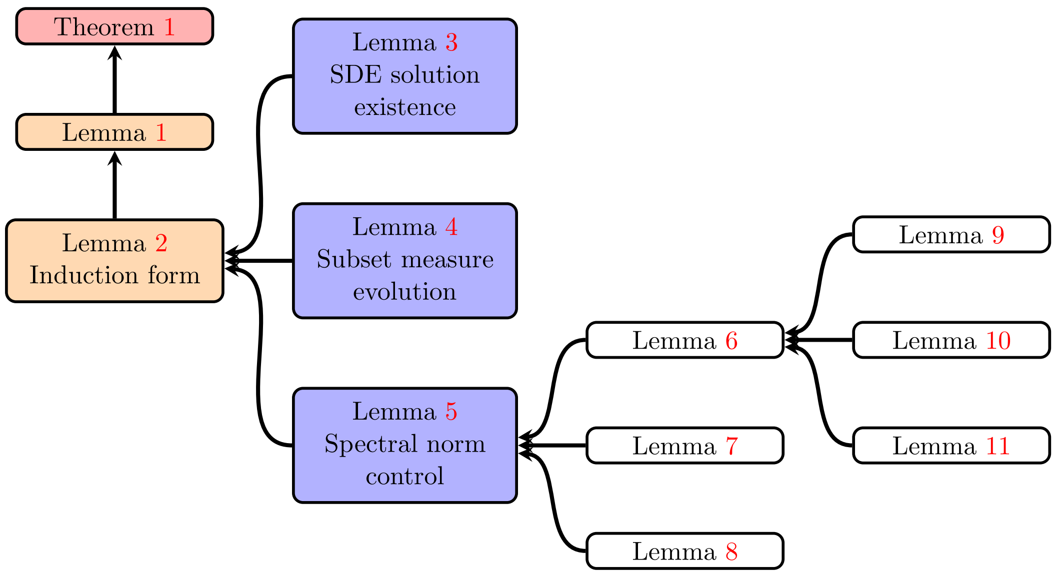


Figure 1: Proof sketch.

the handling of several quantities involved in the stochastic localization scheme. Figure 1 provides a diagram showing the relationship between the main lemmas.

To ensure the existence and the uniqueness of the stochastic localization construction, we first prove a lemma that deals with log-concave densities with compact support. Then we relate back to the main theorem by finding a compact support which contains most of the probability measure for a log-concave density.

Lemma 1. *There exists a universal constant $c$ such that for any log-concave density $p$ in $\mathbb{R}^d$ with compact support and any integer $\ell \geq 1$, we have*

$$\psi(p) \geq \frac{1}{\left[c \cdot \ell\left(\log(d)+1\right)\right]^{\ell/2} d^{16/\ell} \cdot \sqrt{\rho\left(p\right)}}. \tag{6}$$

The proof of Lemma 1 is provided in Section 2.5 after we introduce the intermediate lemmas. The use of the integer $l$ in the lemma indicates that we control the Cheeger isoperimetric coefficient in an iterative fashion. In fact, we prove Lemma 1 by induction over $l$ starting from the known bound in Equation (3). For this, we define the supremum of the product of the isoperimetric coefficient and the square-root of its spectral norm over all log-concave densities in $\mathbb{R}^d$ with compact support:

$$\psi_d = \inf_{\substack{\text{log-concave density } p \text{ in } \mathbb{R}^d \\ \text{with compact support}}} \psi(p)\sqrt{\rho\left(p\right)}. \tag{7}$$

Then we prove the following lemma on the lower bound of $\psi_d$, which serves as the main induction argument.

LEMMA 2. *Suppose that $\psi_k \geq \frac{1}{\alpha k^\beta}$ for all $k \leq d$ for some $0 \leq \beta \leq \frac{1}{2}$ and $\alpha \geq 1$, take $q = \lceil \frac{1}{\beta} \rceil + 1$, there exists a universal constant $c$ such that we have*

$$\psi_d \geq \frac{1}{c \cdot q^{1/2} \alpha \log(d)^{1/2} d^{\beta - \beta/(8q)}}.$$

The proof of Lemma 2 is provided towards the end of this section in Section 2.4. To have a good understanding of how we get there, we start by introducing the stochastic localization scheme introduced by Eldan [8].

**2.1 Eldan's stochastic localization scheme.** Given a log-concave density $p$ in $\mathbb{R}^d$ with covariance matrix $A$, we define the following stochastic differential equation (SDE)

$$\begin{aligned} dc_t &= C_t^{1/2} dW_t + C_t \mu_t dt, \quad c_0 = 0, \\ dB_t &= C_t dt, \quad B_0 = 0, \end{aligned} \tag{8}$$

where $W_t$ is the Wiener process, the matrix $C_t$, the density $p_t$, the mean $\mu_t$ and the covariance $A_t$ are defined as follows

$$C_t = A^{-1}, \tag{9}$$

$$p_t(x) = \frac{e^{c_t^\top x - \frac{1}{2} x^\top B_t x} p(x)}{\int_{\mathbb{R}^d} e^{c_t^\top x - \frac{1}{2} y^\top B_t y} p(y) dy}, \text{for } x \in \mathbb{R}^d, \tag{10}$$

$$\mu_t = \int_{\mathbb{R}^d} x p_t(x) dx, \tag{11}$$

$$A_t = \int_{\mathbb{R}^d} (x - \mu_t)(x - \mu_t) p_t(x) dx. \tag{12}$$

The next lemma shows the existence and the uniqueness of the SDE solution.

LEMMA 3. *Given a density $p$ in $\mathbb{R}^d$ with compact support with covariance $A$ and $A$ is invertible, then the SDE (8) is well defined and it has a unique solution on the time interval $[0, T]$, for any time $T > 0$. Additionally, for any $x \in \mathbb{R}^d$, $p_t(x)$ is a martingale with*

$$dp_t(x) = (x - \mu_t)^\top A^{-1/2} dW_t p_t(x). \tag{13}$$

The proof of Lemma 3 follows from the standard existence and uniqueness theorem of SDE (Theorem 5.2 in Øksendal [21]). The proof is provided in Appendix A.

Before we dive into the proof of Lemma 2, we discuss how the stochastic localization scheme allows us to control the boundary measure of a subset. First, according to the concavity of the isoperimetric profile (Theorem 2.8 in Sternberg and Zumbrun [25] or Theorem 1.8 in Milman [20]), it is sufficient to consider subsets of measure 1/2 in the definition of the isoperimetric coefficient in Equation (2). Second, the density $p_t$ is log-concave and it is more log-concave than the Gaussian density proportional to $e^{-\frac{1}{2} x^\top B_t x}$. It can be shown via the KLS localization lemma [12] that a density which is more log-concave than a Gaussian has an isoperimetric coefficient lower bound that depends on the covariance of the Gaussian (see e.g. Theorem 2.7 in Ledoux [16] or Theorem 4.4 in Cousins and Vempala [7]). Third, given an initial subset $E$ of $\mathbb{R}^d$ with measure $p(E) = \frac{1}{2}$, using the martingale property of $p_t(E)$, we observe that

$$\begin{aligned} p(\partial E) &= \mathbb{E}\left[p_t(\partial E)\right] \\ &\overset{\text{(i)}}{\geq} \mathbb{E}\left[\frac{1}{2}\left\|B_t^{-1}\right\|_2^{-1/2} \min\left(p_t(E), p_t(E^c)\right)\right] \\ &\overset{\text{(ii)}}{\geq} \frac{1}{4} \cdot \frac{1}{2}\left\|B_t^{-1}\right\|_2^{-1/2} \mathbb{P}\left(\frac{1}{4} \leq p_t(E) \leq \frac{3}{4}\right) \\ &= \frac{1}{4}\left\|B_t^{-1}\right\|_2^{-1/2} \mathbb{P}\left(\frac{1}{4} \leq p_t(E) \leq \frac{3}{4}\right) \cdot \min\left\{p(E), p(E^c)\right\}. \end{aligned}$$

Inequality (i) uses the isoperimetric inequality for a log-concave density which is more log-concave than a Gaussian density proportional to $e^{-\frac{1}{2}x^\top B_t x}$ [7, 16]. Inequality (ii) uses the fact that $p_t(E)$ is nonnegative.

Based on the above observation, the high level idea of the proof requires two main steps:

- There exists some time $t > 0$, such that the Gaussian component $\frac{1}{2}x^\top B_t x$ of the density $p_t$ is large enough, so that we can apply the known isoperimetric inequality for densities more log-concave than a Gaussian.
- We need to control the quantity $p_t(E)$ so that the obtained isoperimetric inequality at time $t$ can be related back to that at time 0.

The first step is obvious since our construction explicitly enforces the density $p_t$ to have a Gaussian component $\frac{1}{2}x^\top B_t x$ in Equation (9). Then the remaining question is whether we can run the SDE long enough to make the Gaussian component large enough while still keeping $p_t(E)$ to be the same order as $p(E) = \frac{1}{2}$ with large probability.

### 2.2 Control the evolution of the measure of a subset.

LEMMA 4. *Under the same assumptions of Lemma 3, for any measurable subset $E$ of $\mathbb{R}^d$ with $p(E) = \frac{1}{2}$ and $t > 0$, the solution $p_t$ of the SDE (9) satisfies*

$$\mathbb{P}\left(\frac{1}{4} \leq p_t(E) \leq \frac{3}{4}\right) \geq \frac{9}{10} - \mathbb{P}\left(\int_0^t \left\|A^{-1/2} A_t A^{-1/2}\right\|_2 ds \geq \frac{1}{64}\right).$$

This lemma is proved in Lemma 29 of Lee and Vempala [17]. We provide a proof here for completeness.

*Proof of Lemma 4.* Let $g_t = p_t(E)$. Using Equation (13), we obtain the following derivative of $g_t$

$$dg_t = \int_E (x - \mu_t)^\top A^{-1/2} dW_t p_t(x) dx.$$

Its quadratic variation is

$$\begin{aligned}
d\,[g]_t &= \left\| \int_E A^{-1/2}(x-\mu_t)p_t(x)dx \right\|_2^2 dt \\
&= \max_{\|\xi\|_2 \le 1} \left( \int_E \xi^\top A^{-1/2}(x-\mu_t)p_t(x)dx \right)^2 dt \\
&\le \max_{\|\xi\|_2 \le 1} \left( \int_E \left(\xi^\top A^{-1/2}(x-\mu_t)\right)^2 p_t(x)dx \right) \left( \int_E p_t(x)dx \right) dt \\
&\le \max_{\|\xi\|_2 \le 1} \xi^\top A^{-1/2} A_t A^{-1/2} \xi dt \\
&= \left\| A^{-1/2} A_t A^{-1/2} \right\|_2 dt,
\end{aligned}$$

where the inequality follows from Cauchy–Schwarz inequality. Applying the Dambis, Dubins-Schwarz theorem (see e.g. Revuz and Yor [23] Section V.1 Theorem 1.7), there exists a Wiener process $\tilde{W}_t$ such that $g_t - g_0$ has the same distribution as $\tilde{W}_{[g]_t}$. Since $g_0 = \frac{1}{2}$, we obtain

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{4} \le p_t(E) \le \frac{3}{4}\right) &= \mathbb{P}\left(-\frac{1}{4} \le \tilde{W}_{[g]_t} \le \frac{1}{4}\right) \\
&\ge 1 - \mathbb{P}\left(\max_{0 \le s \le \frac{1}{64}} \left|\tilde{W}_s\right| > \frac{1}{4}\right) - \mathbb{P}([g]_t > \frac{1}{64}) \\
&= 1 - 4\mathbb{P}\left(\tilde{W}_{\frac{1}{64}} > \frac{1}{4}\right) - \mathbb{P}\left([g]_t > \frac{1}{64}\right) \\
&\ge \frac{9}{10} - \mathbb{P}\left(\int_0^t \left\|A^{-1/2} A_t A^{-1/2}\right\|_2 ds > \frac{1}{64}\right),
\end{aligned}$$

where the last inequality follows from the fact that $\mathbb{P}(\xi > 2) < 0.023$ for $\xi$ follows the standard Gaussian distribution. □

**2.3 Control the evolution of the spectral norm.** According to Lemma 4, to control the evolution of the measures of subsets, we need to control the spectral norm of $A^{-1/2}A_tA^{-1/2}$. The following lemma serves the purpose.

LEMMA 5. *In addition to the same assumptions of Lemma 3, if* $\psi_k \ge \frac{1}{\alpha k^\beta}$ *for all* $k \le d$ *for some* $0 < \beta \le \frac{1}{2}$ *and* $\alpha \ge 1$*, then there exists a universal constant* $c$ *such that for* $q = \lceil \frac{1}{\beta} \rceil + 1$, $d \ge 3$ *and* $T_2 = \frac{1}{c \cdot q\alpha^2 \log(d) d^{2\beta - \beta/(4q)}}$*, we have*

$$\mathbb{P}\left(\int_0^{T_2} \left\|A^{-1/2} A_t A^{-1/2}\right\|_2 dt \ge \frac{1}{64}\right) < \frac{4}{10}.$$

Direct control of the largest eigenvalue of $A^{-1/2}A_tA^{-1/2}$ is not trivial, instead we use the potential function $\Gamma_t$ to upper bound the largest eigenvalue. Define

$$Q_t = A^{-1/2} A_t A^{-1/2}$$

$$\Gamma_t = \mathrm{Tr}\left(Q_t^q\right). \tag{14}$$

It is clear that $\Gamma_t^{1/q} \geq \left\|A^{-1/2}A_tA^{-1/2}\right\|_2$. So in order to upper bound $\left\|A^{-1/2}A_tA^{-1/2}\right\|_2$, it is sufficient to upper bound $\Gamma_t^{1/q}$. The advantage of using $\Gamma_t$ is that it is differentiable. We have the following differential for $A_t$ and $\Gamma_t$:

$$dA_t = \int (x-\mu_t)(x-\mu_t)^\top \left((x-\mu_t)^\top A^{-1/2}dW_t\right) p_t(x)dx - A_tA^{-1}A_tdt, \tag{15}$$

$$\begin{aligned} d\Gamma_t = {} & q\int \left(x-\mu_t\right)^\top A^{-1/2}\left(Q_t\right)^{q-1}A^{-1/2}\left(x-\mu_t\right)\left(x-\mu_t\right)^\top A^{-1/2}dW_tp_t(x)dx \\ & - q\,\mathrm{Tr}\left(Q_t^{q+1}\right)dt + \frac{q}{2}\sum_{a=0}^{q-2}\int\int \left(x-\mu_t\right)^\top A^{-1/2}Q_t^aA^{-1/2}\left(y-\mu_t\right) \\ & \cdot \left(x-\mu_t\right)^\top A^{-1/2}Q_t^{q-2-a}A^{-1/2}\left(y-\mu_t\right)\left(x-\mu_t\right)^\top A^{-1}\left(y-\mu_t\right)p_t(x)p_t(y)dxdydt. \end{aligned} \tag{16}$$

Obtaining these differentials uses Itô's formula and the proofs are provided in Appendix A.

The next lemma upper bounds the terms in the potential $\Gamma_t$.

Lemma 6. *Under the same assumptions of Lemma 5, the potential $\Gamma_t$ defined in Equation (14) can be written as follows*

$$d\Gamma_t = v_t^\top dW_t + \delta_t dt,$$

*where $v_t \in \mathbb{R}^d$ and $\delta_t \in \mathbb{R}$ satisfy*

$$\begin{aligned} &\|v_t\|_2 \leq 16q\Gamma_t^{1+1/(2q)}, \text{ and} \\ &\delta_t \leq \min\left\{64q^2\alpha^2\log(d)d^{2\beta-1/q}\Gamma_t^{1+1/q}, \frac{2q^2}{t}\Gamma_t\right\}. \end{aligned}$$

The proof of Lemma 6 is provided in Section 3.1. Remark that bounds similar to the first bound of $\delta_t$ in Lemma 6 have appeared in Lee and Vempala [17], whereas the second bound of $\delta_t$ in Lemma 6 is novel. The second bound of $\delta_t$ also leads to the following Lemma 8 which gives better control of the potential than the previous proof by Lee and Vempala [17] when $t$ is large.

Using the bounds in Lemma 6, we state the two lemmas which control the potential $\Gamma_t$ in two ways.

Lemma 7. *Under the same assumptions of Lemma 6, using the following transformation*

$$\begin{aligned} h : \mathbb{R}_+ &\to \mathbb{R} \\ a &\mapsto -(a+1)^{-1/q} \end{aligned}$$

*we have*

$$\mathbb{P}\left(\max_{t\in[0,T_1]} h(\Gamma_t) \geq -\frac{1}{2}\left(d+1\right)^{-1/q}\right) \leq \exp(-\frac{2}{3}q\log(d)) \leq \frac{3}{10}$$

*where* $T_1 = \frac{1}{32768 q\alpha^2 \log(d) d^{2\beta}}$.

LEMMA 8. *Under the same assumptions of Lemma 6, using the following transformation*

$$\begin{aligned} f : \mathbb{R}_+ &\to \mathbb{R} \\ a &\mapsto a^{1/q} \end{aligned}$$

*we have*

$$\mathbb{E} f(\Gamma_{t_2}) \leq \mathbb{E} f(\Gamma_{t_1}) \left(\frac{t_2}{t_1}\right)^{2q}, \forall t_2 > t_1 > 0.$$

The proofs of Lemma 7 and 8 are provided in Section 3.2.

Now we are ready to prove Lemma 5.

*Proof of Lemma 5.* We take

$$T_1 = \frac{1}{32768 q\alpha^2 \log(d) d^{2\beta}}, \quad T_2 = \frac{d^{\beta/(4q)}}{40} T_1 = \frac{1}{1310720 q\alpha^2 \log(d) d^{2\beta - \beta/(4q)}}.$$

We bound the spectral norm of $A^{-1/2} A_t A^{-1/2}$ in two time intervals via Lemma 7 and Lemma 8. In the first time interval $[0, T_1]$, we have

$$\begin{aligned} \mathbb{P}\left(\int_0^{T_1} \left\| A^{-1/2} A_t A^{-1/2} \right\|_2 \geq \frac{1}{128}\right) &\leq \mathbb{P}\left(\max_{t\in[0,T_1]} \left\| A^{-1/2} A_t A^{-1/2} \right\|_2 \geq \frac{1}{128 T_1}\right) \\ &\overset{\text{(i)}}{\leq} \mathbb{P}\left(\max_{t\in[0,T_1]} \left\| A^{-1/2} A_t A^{-1/2} \right\|_2 \geq 3 d^{1/q}\right) \\ &\overset{\text{(ii)}}{\leq} \mathbb{P}\left(\max_{t\in[0,T_1]} \Gamma_t \geq 3^q d\right) \\ &\overset{\text{(iii)}}{\leq} \mathbb{P}\left(\max_{t\in[0,T_1]} \Gamma_t + 1 \geq 2^q (d+1)\right) \\ &\overset{\text{(iii)}}{=} \mathbb{P}\left(\max_{t\in[0,T_1]} h(\Gamma_t) \geq -\frac{1}{2} (d+1)^{-1/q}\right) \\ &\overset{\text{(iv)}}{\leq} \frac{3}{10}. \end{aligned} \tag{17}$$

Inequality (i) follows from the condition $\beta q \geq 1$. (ii) follows from the fact that $\mathrm{Tr}\,(A^q)^{1/q} \geq \|A\|_2$. (iii) is because $3^q d \geq 2^q (d+1)$ when $q \geq 2$ and $d \geq 1$. $h$ is defined in Lemma 7. (iv) follows from Lemma 7.

In the first time interval, we can also bound the expectation of $\Gamma_{T_1}^{1/q}$. Since the density $p_{T_1}$ is more log-concave than a Gaussian density with covariance matrix $\frac{A}{T_1}$, the covariance matrix of $p_{T_1}$ is upper bounded as follows (see Theorem 4.1 in Brascamp-Lieb [5] or Lemma 5 in Eldan and Lehec [10])

$$A_{T_1} \preceq \frac{A}{T_1}. \tag{18}$$

Consequently, all the eigenvalues of $Q_{T_1}$ are less than $\frac{1}{T_1}$ and $\Gamma_{T_1}$ is upper bounded by $\frac{d}{T_1^q}$. Using the above bound, we can bound the expectation of $\Gamma_{T_1}^{1/q}$ as follows

$$\begin{aligned}\mathbb{E}\left[\Gamma_{T_1}^{1/q}\right] &= \mathbb{E}\left[\mathbb{1}_{\Gamma_{T_1}\geq 3^q d}\Gamma_{T_1}^{1/q} + \mathbb{1}_{\Gamma_{T_1}<3^q d}\Gamma_{T_1}^{1/q}\right] \\ &\overset{\text{(i)}}{\leq} \frac{d^{1/q}}{T_1}\exp\left(-\frac{2}{3}q\log(d)\right) + 3d^{1/q} \\ &\overset{\text{(ii)}}{\leq} 32768 d^{1/q} q\alpha^2 + 4d^{1/q} \\ &\leq 40000 d^{1/q} q\alpha^2. \end{aligned} \tag{19}$$

Inequality (i) follows from Lemma 7, the inequality $3^q d \geq 2^q(d+1)$ (similar to what we did in the last four steps of Equation (17)) and Equation (18). (ii) follows from $q \geq 2$, $\beta \leq 1/2$ and $d^{1/2} \geq \log(d)$ for $d \geq 3$.

In the second time interval, for $t \in [T_1, T_2]$, we have

$$\begin{aligned}\mathbb{E}\left[\left\|A^{-1/2}A_tA^{-1/2}\right\|_2\right] &\leq \mathbb{E}\left[\Gamma_t^{1/q}\right] \\ &\overset{\text{(i)}}{\leq} \mathbb{E}\left[\Gamma_{T_1}^{1/q}\right]\left(\frac{t}{T_1}\right)^{2q} \\ &\overset{\text{(ii)}}{\leq} \mathbb{E}\left[\Gamma_{T_1}^{1/q}\right]\left(\frac{T_2}{T_1}\right)^{2q} \\ &\overset{\text{(iii)}}{\leq} 1000 d^{\beta/2+1/q} q\alpha^2. \end{aligned} \tag{20}$$

Inequality (i) follows from Lemma 8. (ii) is because $t \leq T_2$. (iii) follows from $T_2 = \frac{d^{\beta/(4q)}}{40}T_1$. Using the above bound, we control the spectral norm in the second time interval via Markov's inequality

$$\begin{aligned}\mathbb{P}\left(\int_{T_1}^{T_2}\left\|A^{-1/2}A_tA^{-1/2}\right\|_2 \geq \frac{1}{128}\right) &\overset{\text{(i)}}{\leq} \frac{\mathbb{E}\left[\int_{T_1}^{T_2}\left\|A^{-1/2}A_tA^{-1/2}\right\|_2 dt\right]}{1/128} \\ &\overset{\text{(ii)}}{\leq} T_2\cdot 1000 d^{\beta/2+1/q} q\alpha^2\cdot 128 \\ &\overset{\text{(iii)}}{<} \frac{1}{10}, \end{aligned} \tag{21}$$

where inequality (i) follows from Markov's inequality and (ii) follows from Equation (20). (iii) follows from the definition of $T_2$ and $\frac{\beta}{2}+\frac{1}{q} \leq 2\beta - \beta/(4q)$ when $\beta q \geq 1$ and $q \geq 2$.

Combining the bounds in the first and second time intervals in Equation (17) and (21), we obtain

$$\mathbb{P}\left(\int_0^{T_2} \left\|A^{-1/2}A_tA^{-1/2}\right\|_2 \geq \frac{1}{64}\right) \leq \mathbb{P}\left(\int_0^{T_1} \left\|A^{-1/2}A_tA^{-1/2}\right\|_2 \geq \frac{1}{128}\right) + \mathbb{P}\left(\int_{T_1}^{T_2} \left\|A^{-1/2}A_tA^{-1/2}\right\|_2 \geq \frac{1}{128}\right) \leq \frac{4}{10}. \tag{22}$$

□

**2.4 Proof of Lemma 2.** The proof of Lemma 2 follows the strategy described after Lemma 3. We make the arguments rigorous here. We consider a log-concave density $p$ in $\mathbb{R}^d$ with compact support. Without loss of generality, we can assume that the covariance matrix $A$ of the density $p$ is invertible. Otherwise, the density $p$ is degenerate and we can instead prove the results in a lower dimension.

According to the concavity of the isoperimetric profile (Theorem 2.8 in Sternberg and Zumbrun [25] or Theorem 1.8 in Milman [20]), it is sufficient to consider subsets of measure 1/2 in the definition of isoperimetric coefficient (2). Given an initial subset $E$ of $\mathbb{R}^d$ with $p(E) = \frac{1}{2}$, use the martingale property of $p_{T_2}(E)$, we have

$$\begin{aligned}
p(\partial E) &= \mathbb{E}\left[p_{T_2}(\partial E)\right] \\
&\overset{\text{(i)}}{\geq} \mathbb{E}\left[\frac{1}{2}\left\|B_{T_2}^{-1}\right\|_2^{-1/2} \min\left(p_{T_2}(E), p_{T_2}(E^c)\right)\right] \\
&\overset{\text{(ii)}}{\geq} \frac{1}{4}\cdot\frac{1}{2}\left\|B_{T_2}^{-1}\right\|_2^{-1/2} \mathbb{P}(\frac{1}{4} \leq p_{T_2}(E) \leq \frac{3}{4}) \\
&= \frac{1}{4}\left\|B_{T_2}^{-1}\right\|_2^{-1/2} \mathbb{P}(\frac{1}{4} \leq p_{T_2}(E) \leq \frac{3}{4}) \cdot \min\left\{p(E), p(E^c)\right\} \\
&\overset{\text{(iii)}}{\geq} \frac{1}{8}\left\|B_{T_2}^{-1}\right\|_2^{-1/2} \cdot \min\left\{p(E), p(E^c)\right\} \\
&\overset{\text{(iv)}}{=} \frac{1}{8}T_2^{1/2}\left\|A\right\|_2^{-1/2} \cdot \min\left\{p(E), p(E^c)\right\}.
\end{aligned}$$

Inequality (i) uses the isoperimetric inequality for a log-concave density which is more log-concave than a Gaussian density proportional to $e^{-\frac{1}{2}x^\top B_t x}$ (see e.g. Theorem 2.7 in Ledoux [16] or Theorem 4.4 in Cousins and Vempala [7]). Inequality (ii) follows from the fact that $p_t(E)$ is nonnegative. (iii) follows from Lemma 4 and Lemma 5 (for $d \geq 3$). (iv) follows from the construction that $B_t = tA^{-1}$. We conclude the proof since $T_2$ is taken as $\frac{1}{c\cdot q\alpha^2\log(d)d^{2\beta-\beta/(4q)}}$ with $c$ as a constant. The above proof only works for $d \geq 3$. It is easy to verify that Lemma 2 still holds for the case for $d = 1, 2$ from the original KLS bound in Equation (3). □

**2.5 Proof of Lemma 1.** The proof of Lemma 1 consists of applying Lemma 2 recursively. We define

$$\alpha_1 = 4, \beta_1 = \frac{1}{2}.$$

For $\ell \geq 1$, we define $\alpha_\ell$ and $\beta_\ell$ recursively as follows:

$$\begin{aligned}\alpha_{\ell+1} &= 2c \cdot \alpha_\ell \beta_\ell^{-1/2}, \\ \beta_{\ell+1} &= \beta_\ell - \beta_\ell^2/16, \end{aligned} \tag{23}$$

where $c$ is the constant in Lemma 2. It is not difficult to show by induction that $\alpha_\ell$ and $\beta_\ell$ satisfy

$$\begin{aligned}\frac{1}{\ell+1} \leq \beta_\ell &\leq \frac{16}{\ell} \\ \alpha_\ell &\leq \left(4c^2\ell\right)^{\ell/2}. \end{aligned} \tag{24}$$

We start with a known bound from the original KLS paper [12]

$$\psi_d \geq \frac{1}{\alpha_1 d^{\beta_1}}, \quad \forall d \geq 1.$$

In the induction, suppose that we have

$$\psi_d \geq \frac{1}{\alpha_\ell \left(\log(d)+1\right)^{\ell/2} d^{\beta_\ell}}, \quad \forall d \geq 1.$$

From the above inequality, we obtain for any $1 \leq k \leq d$,

$$\psi_k \geq \frac{1}{\alpha'_\ell k_\ell^\beta},$$

with $\alpha'_\ell = \alpha_\ell \left(\log(d)+1\right)^{\ell/2}$. Using the above lower bounds for $\psi_k$, we can apply Lemma 2. For integer $\ell+1$, we have

$$\begin{aligned}\psi_d &\overset{(i)}{\geq} \frac{1}{c \cdot q^{1/2} \alpha_\ell \left(\log(d)+1\right)^{l/2} \log(d)^{1/2} d^{\beta_\ell - \beta_\ell/(8q)}} \\ &\overset{(ii)}{\geq} \frac{1}{2c \cdot \alpha_\ell \beta_\ell^{-1/2} \left(\log(d)+1\right)^{(l+1)/2} d^{\beta_\ell - \beta_\ell^2/16}} \\ &= \frac{1}{\alpha_{\ell+1} \left(\log(d)+1\right)^{(\ell+1)/2} d^{\beta_{\ell+1}}}\end{aligned}$$

where inequality (i) follows from Lemma 2, inequality (ii) follows from $q \leq \frac{2}{\beta}$ and the last equality follows from the definition of $\alpha_\ell$ and $\beta_\ell$. We conclude Lemma 1 using the $\alpha_\ell$ and $\beta_\ell$ bounds in Equation (24). □

**2.6 Proof of Theorem 1.** To derive Theorem 1 from Lemma 1, it is sufficient to show that for any log-concave density $p$ in $\mathbb{R}^d$, most of its probability measure is on a compact support. Let $\mu$ be the mean of the density $p$. Since $r \mapsto p(\mathbb{B}(\mu, r)^c)$ is an non-increasing function of $r$ with limit 0 at $\infty$, there exists a radius $R > 0$, such that $p(\mathbb{B}(\mu, R)^c) \leq 0.2$. Note that it is possible to get a better bound via e.g. log-concave concentration bounds from Paouris [22], but knowing the existence of such radius $R$ is sufficient for the proof here.

Denote $B = \mathbb{B}(\mu, R)$. Then $p(B^c) \le 0.2$. Let $\varrho$ be the density obtained by truncating $p$ on the ball $B$. Then $\varrho$ is log-concave and it has compact support. For a subset $E \subset \mathbb{R}^d$ of measure such that $p(E) = \frac{1}{2}$, we have

$$\begin{aligned} p(\partial E) &\ge \varrho(\partial E) p(B) \\ &\ge \psi(\varrho) \min\left(\varrho(E), \varrho(E^c)\right) p(B) \\ &= \psi(\varrho) \min\left(p(E \cap B), p(B \cap E^c)\right) \\ &\ge \psi(\varrho) \min\left(p(E) - p(B^c), p(E^c) - p(B^c)\right) \\ &\ge \frac{1}{2}\psi(\varrho) \min\left(p(E), p(E^c)\right). \end{aligned}$$

The last inequality follows because $p(E^c) - p(B^c) \ge 0.5 - 0.2 \ge \frac{1}{4}$. Since it is sufficient to consider subsets of measure 1/2 in the definition of the isoperimetric coefficient [20, 25], we conclude that the isoperimetric coefficient of $p$ is lower bounded by half of that of $\varrho$. Applying Lemma 1 for the isoperimetric coefficient of $\varrho$, we obtain Theorem 1. $\square$

## 3 Proof of auxiliary lemmas

In this section, we prove auxiliary Lemmas 6, 7 and 8.

**3.1 Tensor bounds and proof of Lemma 6.** In this subsection, we prove Lemma 6. Since Lemma 6 involves the third-order moment tensor of a log-concave density, we define the following 3-Tensor for any probability density $p \in \mathbb{R}^d$ with mean $\mu$ to simplify notations.

$$\begin{aligned} \mathcal{T}_p : \quad & \mathbb{R}^{d\times d} \times \mathbb{R}^{d\times d} \times \mathbb{R}^{d\times d} \to \mathbb{R} \\ & (A, B, C) \mapsto \int\int (x-\mu)^\top A (y-\mu) \\ & \cdot (x-\mu)^\top B (y-\mu) \cdot (x-\mu)^\top C (y-\mu) p(x) p(y) dx dy. \end{aligned} \tag{25}$$

For $A, B, C$ three matrices in $\mathbb{R}^{d\times d}$, we can write $\mathcal{T}_p(A, B, C)$ equivalently as

$$\mathcal{T}_p(A, B, C) = \mathbb{E}_{X,Y\sim p}(X-\mu)^\top A(Y-\mu) \cdot (X-\mu)^\top B(Y-\mu) \cdot (X-\mu)^\top C(Y-\mu).$$

Before we prove Lemma 6, we prove the following properties related to the 3-Tensor.

Lemma 9. *Suppose $p$ is a log-concave density with mean $\mu$ and covariance $A$. Then for any positive semi-definite matrices $B$ and $C$, we have*

$$\left\| \int B^{1/2}(x-\mu)(x-\mu)^\top C (x-\mu) p(x) dx \right\|_2 \le 16 \left\| A^{1/2} B A^{1/2} \right\|_2^{1/2} \mathrm{Tr}\left(A^{1/2} C A^{1/2}\right).$$

LEMMA 10. *Suppose that $\psi_k \geq \frac{1}{\alpha k^\beta}$ for all $k \leq d$ for some $0 \leq \beta \leq \frac{1}{2}$ and $\alpha \geq 1$. Suppose $p$ is a log-concave density in $\mathbb{R}^d$ with covariance $A$ and $A$ is invertible. Then for $q \geq \frac{1}{2\beta}$, we have*

$$\mathcal{T}_p(A^{q-2}, \mathbb{I}_d, \mathbb{I}_d) \leq 128\alpha^2 \log(d) d^{2\beta-1/q} \operatorname{Tr}(A^q)^{1+1/q}.$$

LEMMA 11. *Given $\tau > 0$. Suppose $p$ is a log-concave density which is more log-concave than $\mathcal{N}(0, \frac{1}{\tau}\mathbb{I}_d)$. Let $A$ be its covariance matrix. Suppose $A$ is invertible then for $q \geq 3$, we have*

$$\mathcal{T}_p(A^{q-2}, \mathbb{I}_d, \mathbb{I}_d) \leq \frac{4}{\tau} \operatorname{Tr}\left(A^q\right).$$

LEMMA 12. *Suppose $p$ is a log-concave density in $\mathbb{R}^d$. For any $\delta \in [0,1]$, for $A, B, C$ positive semi-definite matrices then*

$$\mathcal{T}_p(B^{1/2}A^\delta B^{1/2}, B^{1/2}A^{1-\delta}B^{1/2}, C) \leq \mathcal{T}_p(B^{1/2}AB^{1/2}, B, C). \tag{26}$$

The proofs of the above lemmas are provided in Section 3.3.

Now we are ready to prove Lemma 6.

*Proof of Lemma 6.* We first prove the bound on $\|v_t\|_2$, where

$$v_t = q \int A^{-1/2}\left(x-\mu_t\right)\left(x-\mu_t\right)^\top A^{-1/2}\left(Q_t\right)^{q-1} A^{-1/2}\left(x-\mu_t\right) p_t(x) dx.$$

Applying Lemma 9 and knowing the covariance of $p_t$ is $A_t$, we obtain

$$\begin{aligned}
\|v_t\|_2 &\leq 16q \left\| A_t^{1/2} A^{-1} A_t^{1/2} \right\|_2^{1/2} \operatorname{Tr}\left( A_t^{1/2} A^{-1/2} Q_t^{q-1} A^{-1/2} A_t^{1/2} \right) \\
&\overset{\text{(i)}}{=} 16q \left\| A_t^{1/2} A^{-1} A_t^{1/2} \right\|_2^{1/2} \operatorname{Tr}\left(Q_t^q\right) \\
&\overset{\text{(ii)}}{=} 16q \left\| Q_t \right\|_2^{1/2} \operatorname{Tr}\left(Q_t^q\right) \\
&\overset{\text{(iii)}}{\leq} 16q \left[\operatorname{Tr}\left(Q_t^q\right)\right]^{1+1/(2q)}.
\end{aligned}$$

Equality (i) uses the definition of $Q_t = A^{-1/2} A_t A^{-1/2}$. Equality (ii) uses the fact that $\left\|MM^\top\right\|_2 = \left\|M^\top M\right\|_2$ for any square matrix $M \in \mathbb{R}^{d\times d}$. Inequality (iii) uses that $\|M\|_2 \leq \operatorname{Tr}\left(M^q\right)^{1/q}$ for any positive semi-definite matrix $M$.

Next, we bound $\delta_t$ in two ways. We can ignore the negative term in $\delta_t$ to obtain the following:

$$\begin{aligned}
\delta_t \leq \frac{q}{2} \sum_{a=0}^{q-2} \int\int & \left(x-\mu_t\right)^\top A^{-1/2} Q_t^a A^{-1/2} \left(y-\mu_t\right) \\
& \cdot \left(x-\mu_t\right)^\top A^{-1/2} Q_t^{q-2-a} A^{-1/2} \left(y-\mu_t\right)\left(x-\mu_t\right)^\top A^{-1}\left(y-\mu_t\right) p_t(x) p_t(y) dx dy \\
= \frac{q}{2} \sum_{a=0}^{q-2} & \mathcal{T}_{\varrho_t}(Q_t^a, Q_t^{q-2-a}, \mathbb{I}_d),
\end{aligned} \tag{27}$$

where $\varrho_t$ is the density of linear-transformed random variable $A^{-1/2}(X-\mu_t)$ for $X$ drawn from $p_t$ and $\mu_t$ is the mean of $p_t$. $\varrho_t$ is still log-concave since any linear transformation of a log-concave density is log-concave (see e.g. Saumard and Wellner [24]). $\varrho_t$ has covariance $A^{-1/2}A_tA^{-1/2}$, which is also $Q_t$. For $a \in \{0, \cdots, q-2\}$, we have

$$\begin{aligned}\mathcal{T}_{\varrho_t}(Q_t^a, Q_t^{q-2-a}, \mathbb{I}_d) &\overset{\text{(i)}}{\leq} \mathcal{T}_{\varrho_t}(Q_t^{q-2}, \mathbb{I}_d, \mathbb{I}_d)\\ &\overset{\text{(ii)}}{\leq} 128\alpha^2 \log(d) d^{2\beta-1/q} \left[\mathrm{Tr}\left(Q_t^q\right)\right]^{1+1/q}.\end{aligned}$$

Inequality (i) follows from Lemma 12. Inequality (ii) follows from Lemma 10. Since there are $q-1$ terms in the sum, we conclude the first part of the bound for $\delta_t$.

On the other hand, since $p_t$ is more log-concave than the Gaussian density proportional to $e^{-\frac{t}{2}(x-\mu_t)^\top A^{-1}(x-\mu_t)}$, $\varrho_t$ is more log-concave than the Gaussian density proportional to $e^{-\frac{t}{2}x^\top x}$. Applying Lemma 12 and Lemma 11 to each term in Equation (27), we obtain

$$\begin{aligned}\delta_t &\leq \frac{q^2}{2}\mathcal{T}_{\varrho_t}(Q_t^{q-2}, \mathbb{I}_d, \mathbb{I}_d)\\ &\leq \frac{2q^2}{t}\,\mathrm{Tr}\left(Q_t^q\right).\end{aligned}$$

This concludes the second part of the bound for $\delta_t$. □

**3.2 Control of the potential in two time intervals.** In this subsection, we prove Lemma 7 and Lemma 8.

*Proof of Lemma 7.* The function $h$ has the following derivatives

$$\frac{dh}{da} = \frac{1}{q}(a+1)^{-1/q-1}, \quad \frac{d^2h}{da^2} = -\frac{q+1}{q^2}(a+1)^{-1/q-2}.$$

Using Itô's formula, we obtain

$$\begin{aligned}dh(\Gamma_t) &= \left.\frac{dh}{da}\right|_{\Gamma_t} d\Gamma_t + \frac{1}{2}\left.\frac{d^2h}{da^2}\right|_{\Gamma_t} d\left[\Gamma\right]_t\\ &= \frac{1}{q(\Gamma_t+1)^{1/q+1}}d\Gamma_t - \frac{1}{2}\frac{q+1}{q^2(\Gamma_t+1)^{1/q+2}}\left\|v_t\right\|_2^2 dt\\ &\leq \frac{1}{q(\Gamma_t+1)^{1/q+1}}d\Gamma_t\\ &\overset{\text{(i)}}{\leq} 64q\alpha^2\log(d)d^{2\beta-1/q}dt + \frac{v_t^\top dW_t}{q\left(\Gamma_t+1\right)^{1/q+1}},\end{aligned}$$

where inequality (i) plugs in the bounds in Lemma 6.

Define a martingale $Y_t$ such that

$$dY_t = \frac{v_t^\top dW_t}{q\left(\Gamma_t+1\right)^{1/q+1}},$$

with $Y_0 = 0$. According to the $\|v_t\|_2$ upper bound in Lemma 6, we have

$$\left\| \frac{1}{q\left(\Gamma_t+1\right)^{1+1/q}} v_t \right\|_2^2 \leq 256.$$

Hence the martingale $Y_t$ is well-defined. According to the Dambis, Dubins-Schwarz theorem (see e.g. Revuz and Yor [23] Section V.1 Theorem 1.7), there exits a Wiener process $\tilde{W}_t$ such that $Y_t$ has the same distribution as $\tilde{W}_{[Y]_t}$. Then we have for any $\gamma > 0$,

$$\mathbb{P}\left(\max_{t\in[0,T]} Y_t \geq \gamma\right) \leq \mathbb{P}\left(\max_{t\in[0,T]} \tilde{W}_{256t} \geq \gamma\right) \leq \exp\left(-\frac{\gamma^2}{512T}\right). \tag{28}$$

Set $T = \frac{1}{32768 q\alpha^2 \log(d) d^{2\beta}}$ and $\Psi = \frac{1}{2}\left(d+1\right)^{-1/q}$. Observe that $\Gamma_0 = d$ and as a result $h(\Gamma_0) = -\left(d+1\right)^{-1/q}$. Then we have

$$\begin{aligned}
\mathbb{P}\left(\max_{t\in[0,T]} h(\Gamma_t) \geq -\Psi\right) &\leq \mathbb{P}\left(\max\right)_{t\in[0,T]} Y_t \geq -\Psi + \left(d+1\right)^{-1/q} \\
&\quad - \int_0^T 64 q\alpha^2 \log(d) d^{2\beta-1/q} dt \\
&\overset{\text{(i)}}{\leq} \mathbb{P}\left(\max_{t\in[0,T]} Y_t \geq \frac{\Psi}{4}\right) \\
&\overset{\text{(ii)}}{\leq} \exp\left(-\frac{\Psi^2}{8192T}\right) \\
&\overset{\text{(iii)}}{\leq} \exp\left(-\frac{2}{3} q\alpha^2 \log(d) d^{2\beta-2/q}\right) \\
&\overset{\text{(iv)}}{<} \frac{3}{10}.
\end{aligned}$$

Inequality (i) follows from the choice of $T$. (ii) uses Equation (28). (iii) follows by plugging in $\Psi = \frac{1}{2}\left(d+1\right)^{-1/q}$ and $3^q d^2 \geq 2^q (d+1)^2$. (iv) follows from $\beta q \geq 1$, $d \geq 3$, $q \geq 2$ and $3^{-4/3} < 0.3$. □

*Proof of Lemma 8.* The function $f$ has the following derivatives

$$\frac{df(a)}{da} = \frac{1}{q} a^{1/q-1}, \frac{d^2 f(a,t)}{da^2} = -\frac{q-1}{q^2} a^{1/q-2}.$$

Using Itô's formula, we obtain

$$\begin{aligned}
df\left(\Gamma_t\right) &= \left.\frac{df}{da}\right|_{\Gamma_t} d\Gamma_t + \frac{1}{2} \left.\frac{d^2 f}{d^2 a}\right|_{\Gamma_t} d\left[\Gamma\right]_t \\
&= \frac{1}{q} \Gamma_t^{1/q-1} \left(v_t^\top dW_t + \delta_t dt\right) - \frac{q-1}{2q^2} \Gamma_t^{1/q-2} \left\|v_t\right\|_2^2 dt.
\end{aligned}$$

Using the bounds in Lemma 6 and the martingale property of the term $\frac{1}{q}\Gamma_t^{1/q-1}v_t^\top dW_t$, we obtain

$$d\mathbb{E}f(\Gamma_t) \le \frac{2q}{t}\mathbb{E}f(\Gamma_t)dt.$$

Solving the above differential equation, we obtain

$$\mathbb{E}f(\Gamma_{t_2}) \le \mathbb{E}f(\Gamma_{t_1})\left(\frac{t_2}{t_1}\right)^{2q}, \forall t_2 > t_1 > 0.$$

□

**3.3 Proof of tensor bounds.** In this subsection, we prove Lemmas 9, 10, 11 and 12.

*Proof of Lemma 9*. Since $C$ is positive semi-definite, we can write its eigenvalue decomposition as follows $C = \sum_{i=1}^d \lambda_i v_i v_i^\top$, with $\lambda_i \ge 0$. Then,

$$\begin{aligned}
&\left\|\int B^{1/2}(x-\mu)(x-\mu)^\top C(x-\mu)p(x)dx\right\|_2 \\
&= \left\|\sum_{i=1}^d \int B^{1/2}(x-\mu)\lambda_i\left((x-\mu)^\top v_i\right)^2 p(x)dx\right\|_2 \\
&\overset{\text{(i)}}{\le} \sum_{i=1}^d \lambda_i \left\|\int B^{1/2}(x-\mu)\left((x-\mu)^\top v_i\right)^2 p(x)dx\right\|_2 \\
&= \sum_{i=1}^d \lambda_i \max_{\|\xi\|_2\le 1}\int \xi^\top B^{1/2}(x-\mu)\left((x-\mu)^\top v_i\right)^2 p(x)dx \\
&\overset{\text{(ii)}}{\le} \sum_{i=1}^d \lambda_i \max_{\|\xi\|_2\le 1}\left(\int\left(\xi^\top B^{1/2}(x-\mu)\right)^2 p(x)dx\right)^{1/2}\left(\int\left((x-\mu)^\top v_i\right)^4 p(x)dx\right)^{1/2} \\
&\overset{\text{(iii)}}{\le} 16\sum_{i=1}^d \lambda_i \max_{\|\xi\|_2\le 1}\left(\int\left(\xi^\top B^{1/2}(x-\mu)\right)^2 p(x)dx\right)^{1/2}\left(\int\left((x-\mu)^\top v_i\right)^2 p(x)dx\right) \\
&= 16\left\|B^{1/2}AB^{1/2}\right\|_2^{1/2}\operatorname{Tr}\left(A^{1/2}CA^{1/2}\right).
\end{aligned}$$

Inequality (i) follows from triangular inequality. (ii) follows from Cauchy–Schwarz inequality. (iii) follows from the statement below, which upper bounds the fourth moment of a log-concave density via its second moment. □

For any log-concave density $\nu$ and any vector $\theta \in \mathbb{R}^d$, we have

$$\left(\int\left((x-\mu_\nu)^\top\theta\right)^a \nu(x)dx\right)^{1/a} \le 2\frac{a}{b}\left(\int\left((x-\mu_\nu)^\top\theta\right)^b \nu(x)dx\right)^{1/b} \tag{29}$$

for $a \geq b > 0$, where $\mu_\nu$ is the mean of $\nu$. Equation (29) is proved e.g. in Corollary 5.7 of Guédon et al. [11] and the exact constant is provided in Proposition 3.8 of Latała and Wojtaszczyk [15].

In order to prove Lemma 10, we need to introduce one additional lemma as follows.

LEMMA 13. *Suppose that $\psi_k \geq \frac{1}{\alpha k^\beta}$ for all $k \leq d$ for some $0 < \beta \leq \frac{1}{2}$ and $\alpha \geq 1$. For an isotropic log-concave density $p$ in $\mathbb{R}^d$ and a unit vector $v \in \mathbb{R}^d$, define $\Delta = \mathbb{E}_{X\sim p}\left(X^\top v\right) \cdot XX^\top$, then we have*

1. *For any orthogonal projection matrix $P \in \mathbb{R}^{d\times d}$ with rank $r$, we have*
$$\mathrm{Tr}\left(\Delta P \Delta\right) \leq 16\psi_{\min(2r,d)}^{-2}.$$
2. *For any positive semi-definite matrix $A$, we have*
$$\mathrm{Tr}\left(\Delta A \Delta\right) \leq 128\alpha^2 \log(d)\left(\mathrm{Tr}\left(A^{1/(2\beta)}\right)\right)^{2\beta}.$$

This lemma was proved in Lemma 41 in an older version (arXiv version 2) of Lee and Vempala [17]. The main proof idea for the first part of Lemma 13 appeared in Eldan [8] (Lemma 6). we provide a proof here for completeness.

*Proof of Lemma 13*. For the first part, we have
$$\mathrm{Tr}\left(\Delta P \Delta\right) = \mathbb{E}_{X\sim p} X^\top \Delta P X \cdot X^\top v.$$
Since $\mathbb{E}_{X\sim p}X^\top v = 0$, we can subtract the mean of the first term $X^\top \Delta P X$ without changing the value of $\mathrm{Tr}\left(\Delta P \Delta\right)$. Then
$$\begin{aligned}
\mathrm{Tr}\left(\Delta P \Delta\right) &= \mathbb{E}_{X\sim p}\left[\left(X^\top \Delta P X - \mathbb{E}_{Y\sim p} Y^\top \Delta P Y\right)\cdot X^\top v\right] \\
&\overset{\text{(i)}}{\leq} \left(\mathbb{E}_{X\sim p}(X^\top v)^2\right)^{1/2}\left(\mathrm{Var}_{X\sim p}\left(X^\top \Delta P X\right)\right)^{1/2} \\
&\overset{\text{(ii)}}{\leq} 2\psi_{\min(2r,d)}^{-1}\left(\mathbb{E}_{X\sim p}\left\|\Delta P X + P^\top \Delta^\top X\right\|_2^2\right)^{1/2} \\
&\leq 4\psi_{\min(2r,d)}^{-1}\left(\mathrm{Tr}\left(\Delta P \Delta\right)\right)^{1/2}.
\end{aligned}$$
Inequality (i) follows from the Cauchy–Schwarz inequality. Inequality (ii) follows from the fact that $\mathbb{E}_{X\sim p}(X^\top v)^2 = 1$ as $p$ is isotropic and that the inverse Poincaré constant is upper bounded by twice of inverse of the squared isoperimetric coefficient (also known as Cheeger's inequality [6, 19] or Theorem 1.1 in Milman [20]). The matrix $\Delta P + P^\top \Delta$ has rank at most $\min(2r, d)$. Rearranging the terms in the above equation, we conclude the first part of Lemma 13.

For the second part, we write the matrix $A$ in its eigenvalue decomposition and group the terms by eigenvalues. We have
$$A = \sum_{i=1}^d \lambda_i v_i v_i^\top = \sum_{j=1}^J A_j + B,$$

where $A_i$ has eigenvalues between the interval $(\|A\|_2 e^{i-1}/d, \|A\|_2 e^i/d]$ and $B$ has eigenvalues smaller than or equal to $\|A\|_2/d$. Because the intervals have right bounds increasing exponentially, we have $J = \lceil \log(d) \rceil$. Let $P_i$ be the orthogonal projection matrix formed by the eigenvectors in $A_i$. Then we have

$$\begin{aligned} \operatorname{Tr}(\Delta A_i \Delta) \leq \|A_i\|_2 \operatorname{Tr}(\Delta P_i \Delta) &\overset{\text{(i)}}{\leq} 16 \|A_i\|_2 \psi^{-2}_{\min(2\mathrm{rank}(A_i), d)} \overset{\text{(ii)}}{} \\ &\leq 16\alpha^2 \|A_i\|_2 \cdot (2\mathrm{rank}(A_i))^{2\beta}, \end{aligned} \tag{30}$$

where inequality (i) follows from the first part of Lemma 13 and inequality (ii) follows from the hypothesis of Lemma 13. Similarly for matrix $B$, we have

$$\operatorname{Tr}(\Delta B \Delta) \overset{\text{(i)}}{\leq} 16\alpha^2 \|B\|_2 (2\mathrm{rank}(B))^{2\beta} \overset{\text{(ii)}}{\leq} 32\alpha^2 \|A\|_2, \tag{31}$$

where inequality (i) follows from the hypothesis of Lemma 13 and inequality (ii) follows from the fact that $\|B\|_2 \leq \|A\|_2/d$ and $2\beta \leq 1$. Putting the bounds (30) and (31) together, we have

$$\begin{aligned} \operatorname{Tr}(\Delta A \Delta) &= \sum_{j=1}^{J} \operatorname{Tr}(\Delta A_j \Delta) + \operatorname{Tr}(\Delta B \Delta) \\ &\leq 16\alpha^2 \left( \sum_{j=1}^{J} \|A_j\|_2 \cdot (2\mathrm{rank}(A_j))^{2\beta} + 2\|A\|_2 \right) \\ &\overset{\text{(i)}}{\leq} 16\alpha^2 \left[ \left( \sum_{j=1}^{J} \|A_j\|_2^{1/(2\beta)} \cdot (2\mathrm{rank}(A_j)) \right)^{2\beta} \cdot (J)^{1-2\beta} + 2\|A\|_2 \right] \\ &\overset{\text{(ii)}}{\leq} 16\alpha^2 \left[ \left( 2e \operatorname{Tr}\left(A^{1/(2\beta)}\right)\right)^{2\beta} \cdot (J)^{1-2\beta} + 2\|A\|_2 \right] \\ &\leq 128\alpha^2 \log(d) \left( \operatorname{Tr}\left(A^{1/(2\beta)}\right)\right)^{2\beta}. \end{aligned}$$

Inequality (i) follows from Holder's inequality and inequality (ii) follows from the fact that $\|A_j\|_2^{1/2\beta} \mathrm{rank}(A_j) \leq e \operatorname{Tr}\left(A_j^{1/2\beta}\right)$ due to the construction of $A_j$. This concludes the second part of Lemma 13. $\square$

*Proof of Lemma 10.* Let $\mu$ be the mean of $p$. First, for $X$ a random vector in $\mathbb{R}^d$ drawn from $p$, we define the standardized random variable $A^{-1/2}(X-\mu)$ and its density $\varrho$. $\varrho$ is an isotropic log-concave density. Then through a change of variable, we have

$$\begin{aligned} &\mathcal{T}_p\left(A^{q-2}, \mathbb{I}_d, \mathbb{I}_d\right) \\ &= \int\int (x-\mu)^\top A^{q-2}(y-\mu) \cdot (x-\mu)^\top (y-\mu) \cdot (x-\mu)^\top (y-\mu) p(x)p(y) dx dy \end{aligned}$$

$$
\begin{aligned}
&= \int\int \left(x^\top A^{q-1} y\right)(x^\top A y)(x^\top A y)\varrho(x)\varrho(y)dxdy \\
&\leq \int\int \left(x^\top A^{q} y\right)(x^\top A y)(x^\top y)\varrho(x)\varrho(y)dxdy \\
&= \mathcal{T}_\varrho\left(A^q, A, \mathbb{I}_d\right),
\end{aligned}
$$

where the last inequality follows from Lemma 12. $A^q$ is positive semi-definite and we write down its eigenvalue decomposition $A^q = \sum_{i=1}^d \lambda_i v_i v_i^\top$ with $\lambda_i \geq 0$. Since $\varrho$ is isotropic, we can rewrite the 3-Tensor into a summation form and apply Lemma 13.

$$
\begin{aligned}
&\mathcal{T}_\varrho\left(A^q, A, \mathbb{I}_d\right) \\
&\quad = \int\int \left(x^\top A^q y\right)\left(x^\top A y\right)\left(x^\top y\right)\varrho(x)\varrho(y)dxdy \\
&\quad = \sum_{i=1}^d \lambda_i \int\int \left(x^\top v_i\right)\left(y^\top v_i\right)\left(x^\top A y\right)\left(x^\top y\right)\varrho(x)\varrho(y)dxdy \\
&\quad = \sum_{i=1}^d \lambda_i \operatorname{Tr}\left(\Delta_i A \Delta_i\right) \\
&\quad \overset{\text{(i)}}{\leq} 128\alpha^2 \log(d)\left(\operatorname{Tr}(A^{1/2\beta})\right)^{2\beta}\left(\sum_{i=1}^d \lambda_i\right) \\
&\quad = 128\alpha^2 \log(d)\left(\operatorname{Tr}(A^{1/2\beta})\right)^{2\beta}\operatorname{Tr}(A^q) \\
&\quad \overset{\text{(ii)}}{\leq} 128\alpha^2 \log(d)\operatorname{Tr}(A^q)\left[\operatorname{Tr}\left(A^q\right)^{1/(2\beta q)}(d)^{1-1/(2\beta q)}\right]^{2\beta} \\
&\quad = 128\alpha^2 \log(d) d^{2\beta-1/q}\operatorname{Tr}(A^q)^{1+1/q},
\end{aligned}
$$

where we define $\Delta_i = \int (x^\top v_i) x x^\top \varrho(x)dx$, inequality (i) follows from Lemma 13 and that $\varrho$ is isotropic, inequality (ii) follows from Cauchy–Schwarz inequality and the assumption that $q \geq \frac{1}{2\beta}$. □

*Proof of Lemma 11.* Without loss of generality, we can assume that the density $p$ has mean 0. Its covariance matrix $A$ is positive semi-definite and invertible. We can write down its eigenvalue decomposition as follows $A = \sum_{i=1}^d \lambda_i v_i v_i^\top$ with $\lambda_i > 0$ and $v_i$ are eigenvectors with norm 1. Then $A^q$ has an eigenvalue decomposition with the same eigenvectors $A^q = \sum_{i=1}^d \lambda_i^q v_i v_i^\top$. Define $\Delta_i = \mathbb{E}_{X\sim p}(X^\top A^{-1/2} v_i) X X^\top$, then

$$
\begin{aligned}
\mathcal{T}_p\left(A^{q-2}, \mathbb{I}_d, \mathbb{I}_d\right) &= \mathbb{E}_{X,Y\sim p}\left(X^\top A^{q-2} Y\right)(X^\top Y)(X^\top Y) \\
&= \sum_{i=1}^d \lambda_i^{q-1}\operatorname{Tr}\left(\Delta_i \Delta_i\right). \qquad (32)
\end{aligned}
$$

Next we bound the terms $\mathrm{Tr}\left(\Delta_i\Delta_i\right)$. We have

$$\begin{aligned}
\mathrm{Tr}\left(\Delta_i\Delta_i\right) &= \mathbb{E}_{X\sim p}\left(X^\top A^{-1/2}v_i\right)X^\top\Delta_i X\\
&\overset{\text{(i)}}{=} \mathbb{E}_{X\sim p}\left(X^\top A^{-1/2}v_i\right)\left(X^\top\Delta_i X - \mathbb{E}_{Y\sim p}\left[Y^\top\Delta_i Y\right]\right)\\
&\overset{\text{(ii)}}{\le} \left(\mathbb{E}_{X\sim p}\left(X^\top A^{-1/2}v_i\right)^2\right)^{1/2}\left(\mathrm{Var}\left(X^\top\Delta_i X\right)\right)^{1/2}\\
&\overset{\text{(iii)}}{=} \left(\mathrm{Var}_{X\sim p}\left(X^\top\Delta_i X\right)\right)^{1/2}\\
&\overset{\text{(iv)}}{\le} \left(\mathbb{E}_{X\sim p}\frac{1}{\tau}\left\|\Delta_i X + \Delta_i X\right\|_2^2\right)^{1/2}\\
&\overset{\text{(v)}}{\le} \left(\frac{4}{\tau}\,\mathrm{Tr}\left(A\Delta_i\Delta_i\right)\right)^{1/2}.
\end{aligned}$$

Equality (i) is because $\mathbb{E}_{X\sim p}X = 0$. Inequality (ii) follows from Cauchy–Schwarz inequality. Equality (iii) follows from the definition of the covariance matrix $\mathbb{E}_{X\sim p}XX^\top = A$. Inequality (iv) follows from the Brascamp-Lieb inequality (or Hessian Poincaré, see Theorem 4.1 in Brascamp and Lieb [5]) together with the assumption that $p$ is more log-concave than $\mathcal{N}(0, \frac{1}{\tau}\mathbb{I}_d)$.

Plugging the bounds of the terms $\mathrm{Tr}\left(\Delta_i\Delta_i\right)$ into Equation (32), we obtain

$$\begin{aligned}
\mathcal{T}_p\left(A^{q-2}, \mathbb{I}_d, \mathbb{I}_d\right) &= \sum_{i=1}^d \lambda_i^{q-1}\,\mathrm{Tr}\left(\Delta_i\Delta_i\right)\\
&\le \sum_{i=1}^d \lambda_i^{q-1}\left(\frac{4}{\tau}\,\mathrm{Tr}\left(A\Delta_i\Delta_i\right)\right)^{1/2}\\
&\overset{\text{(i)}}{\le} \frac{2}{\tau^{1/2}}\left(\sum_{i=1}^d \lambda_i^q\right)^{1/2}\left(\sum_{i=1}^d \lambda_i^{q-2}\,\mathrm{Tr}\left(A\Delta_i\Delta_i\right)\right)^{1/2}\\
&= \frac{2}{\tau^{1/2}}\left(\mathrm{Tr}\left(A^q\right)\right)^{1/2}\left(\mathbb{E}_{X,Y\sim p}\left(X^\top A^{q-3}Y\right)(X^\top AY)(X^\top Y)\right)^{1/2}\\
&\overset{\text{(ii)}}{\le} \frac{2}{\tau^{1/2}}\left(\mathrm{Tr}\left(A^q\right)\right)^{1/2}\left(\mathbb{E}_{X,Y\sim p}\left(X^\top A^{q-2}Y\right)(X^\top Y)(X^\top Y)\right)^{1/2}\\
&= \frac{2}{\tau^{1/2}}\left(\mathrm{Tr}\left(A^q\right)\right)^{1/2}\left[\mathcal{T}_p\left(A^{q-2}, \mathbb{I}_d, \mathbb{I}_d\right)\right]^{1/2}.
\end{aligned}$$

Inequality (i) follows from Cauchy–Schwarz inequality. For $q \ge 3$, inequality (ii) follows from Lemma 12. From the above equation, after rearranging the terms, we obtain

$$\mathcal{T}_p\left(A^{q-2}, \mathbb{I}_d, \mathbb{I}_d\right) \le \frac{4}{\tau}\,\mathrm{Tr}\left(A^q\right).$$

$\square$

*Proof of Lemma 12*. This lemma is proved in Lemma 43 in an older version (arXiv version 2) of Lee and Vempala [17], we provide a proof here for completeness.

Without loss of generality, we can assume that the density $p$ has mean 0. For $i \in \{1, \cdots, d\}$, we define $\Delta_i = \mathbb{E}_{X\sim p} B^{1/2} X X^\top B^{1/2} X^\top C^{1/2} e_i$ where $e_i \in \mathbb{R}^d$ is the vector with $i$th coordinate 1 and 0 elsewhere. We have $\sum_{i=1}^d e_i e_i^\top = \mathbb{I}_d$. We can rewrite the tensor on the left hand side as a sum of traces.

$$\begin{aligned}
&\mathcal{T}_p(B^{1/2}A^\delta B^{1/2}, B^{1/2}A^{1-\delta}B^{1/2}, C)\\
&\quad = \mathbb{E}_{X,Y\sim p} X^\top B^{1/2} A^\delta B^{1/2} Y \cdot X^\top B^{1/2} A^{1-\delta} B^{1/2} Y \cdot X^\top C Y\\
&\quad = \sum_{i=1}^d \mathbb{E}_{X,Y\sim p} X^\top B^{1/2} A^\delta B^{1/2} Y \cdot X^\top B^{1/2} A^{1-\delta} B^{1/2} Y \cdot X^\top C^{1/2} e_i \cdot Y^\top C^{1/2} e_i\\
&\quad = \sum_{i=1}^d \mathrm{Tr}\left(A^\delta \Delta_i A^{1-\delta} \Delta_i\right). \qquad (33)
\end{aligned}$$

For any symmetric matrix $F$, a positive-semidefinite matrix $G$ and $\delta \in [0, 1]$, we have

$$\mathrm{Tr}\left(G^\delta F G^{1-\delta} F\right) \le \mathrm{Tr}\left(G F^2\right). \qquad (34)$$

Applying the above trace inequality (34) that we prove later for completeness (see also Lemma 2.1 in Zhu et al. [1]), we obtain

$$\mathrm{Tr}\left(A^\delta \Delta_i A^{1-\delta} \Delta_i\right) \le \mathrm{Tr}\left(A \Delta_i \Delta_i\right).$$

Writing the sum of traces in Equation (33) back to the 3-Tensor form, we conclude Lemma 12.

It remains to prove the trace inequality in Equation (34). Without loss of generality, we can assume $G$ is diagonal. Hence, we have

$$\begin{aligned}
\mathrm{Tr}\left(G^\delta F G^{1-\delta} F\right) &= \sum_{i=1}^d \sum_{j=1}^d G_{ii}^\delta G_{jj}^{1-\delta} F_{ij}^2\\
&\le \sum_{i=1}^d \sum_{j=1}^d \left(\delta G_{ii} + (1-\delta) G_{jj}\right) F_{ij}^2\\
&= \delta \sum_{i=1}^d \sum_{j=1}^d G_{ii} F_{ij}^2 + (1-\delta) \sum_{i=1}^d \sum_{j=1}^d G_{jj} F_{ij}^2\\
&= \mathrm{Tr}\left(G F^2\right),
\end{aligned}$$

where the inequality follows from Jensen's inequality and the fact that the logarithm function is concave (or the inequality of arithmetic and geometric means). □

## Acknowledgments

Yuansi Chen has received funding from the European Research Council under the Grant Agreement No 786461 (CausalStats - ERC-2017-ADG). We acknowledge scientific interaction and exchange at "ETH Foundations of Data Science". We thank Peter Bühlmann and Bin Yu for their continuous support and encouragement. We thank Afonso Bandeira, Raaz Dwivedi, Ronen Eldan, Yin Tat Lee and Martin Wainwright for helpful discussions. We thank Bo'az Klartag and Joseph Lehec for pointing out a mistake in the previous revision. We also thank anonymous reviewers for their careful reading of our manuscript and their suggestions on presentation and writing.

**Funding** Open Access funding provided by ETH Zurich



## A Proof of Lemma 3 and derivatives

In this section, we first prove the existence and uniqueness of the SDE solution in Lemma 3 and then derive the derivatives of $p_t$, $A_t$ and $\Gamma_t$ in Equation (13), Equation (15) and (16) using Itô's calculus. Similar results are also proved in Eldan [8] and Lee and Vempala [17] since a similar stochastic localization is used. We provide a proof here for completeness.

*Proof of Lemma 3.* We can rewrite the stochastic differential equation (8) as follows to make the dependency clear:

$$\begin{aligned} dc_t &= A^{-1/2} dW_t + A^{-1} \mu\left(c_t, B_t\right) dt \\ dB_t &= A^{-1} dt, \end{aligned}$$

where

$$\begin{aligned} \mu(c, B) &= \int x \varrho(c, B, x) dx, \\ \varrho(c, B, x) &= \frac{e^{c^\top x - \frac{1}{2} x^\top B x} p(x)}{\int_{\mathbb{R}^d} e^{c^\top x - \frac{1}{2} y^\top B y} p(y) dy}. \end{aligned}$$

Since $p$ has a compact support, given $x \in \mathbb{R}^d$, $\varrho(\cdot,\cdot,x)$ as a function of $(c,B)$ is Lipschitz in $c$ and $B$. Similarly, $\mu$ is also Lipschitz in $c$ and $B$. Consequently, $A^{-1/2}$, $A^{-1}\mu(c_t,B_t)$ and $A^{-1}$ are all bounded and Lipschitz on $c_t$ and $B_t$ on the compact support. Applying the existence and uniqueness theorem of SDE solutions (Theorem 5.2 in Øksendal [21]), we show that the SDE solution exists and is unique on the time interval $[0,T]$ for any $T>0$.

Next, we derive the derivative of $p_t$. Define

$$\begin{aligned} G_t(x) &= e^{c_t^\top x - \frac{1}{2}x^\top B_t x} p(x), \\ V_t &= \int G_t(x)dx. \end{aligned}$$

Then $p_t(x)$ can be written as $\frac{G_t(x)}{V_t}$. Let $S_t(x)$ denote the quadratic variation of the process $c_t^\top x$. We have

$$dS_t(x) = x^\top A^{-1} x dt.$$

Using Itô's formula, we have

$$\begin{aligned} dG_t(x) &= \left( x^\top (dc_t) - \frac{1}{2} x^\top dB_t x + \frac{1}{2} dS_t \right) G_t(x) \\ &= \left( x^\top A^{-1/2} dW_t + x^\top A^{-1} \mu_t dt \right) G_t(x), \\ dV_t &= \int dG_t(x)dx = V_t \left( \mu_t^\top A^{-1/2} dW_t + \mu_t^\top A^{-1} \mu_t dt \right). \end{aligned}$$

Using Itô's formula on the inverse of $V_t$, we have

$$\begin{aligned} dV_t^{-1} &= -\frac{dV_t}{V_t^2} + \frac{d\,[V]_t}{V_t^3} \\ &= -V_t^{-1} \left[ \mu_t^\top A^{-1/2} dW_t + \mu_t^\top A^{-1} \mu_t dt \right] + V_t^{-1} \mu_t^\top A^{-1} \mu_t dt \\ &= -V_t^{-1} \mu_t^\top A^{-1/2} dW_t. \end{aligned}$$

Using Itô's formula on $p_t$, with the above derivatives, we obtain

$$\begin{aligned} dp_t(x) &= d\left( V_t^{-1} G_t(x) \right) \\ &= \left( G_t(x) dV_t^{-1} + V_t^{-1} dG_t(x) + d\left[ V^{-1}, G(x) \right]_t \right) \\ &= (x - \mu_t)^\top A^{-1/2} dW_t p_t(x). \end{aligned}$$

Then we derive the derivative of $A_t$. By the definition of $A_t$, we have

$$A_t = \int (x-\mu_t)(x-\mu_t)^\top p_t(x) dx,$$

where $\mu_t = \int_{\mathbb{R}^d} xp_t(x)dx$. Using Itô's formula on $\mu_t$, we obtain

$$\begin{aligned} d\mu_t &= \int xdp_t(x)dx \\ &= \int x(x-\mu_t)^\top A^{-1/2} dW_t p_t(x)dx \\ &= \int (x-\mu_t)(x-\mu_t)^\top A^{-1/2} dW_t dx \\ &= A_t A^{-1/2} dW_t. \end{aligned}$$

Using Itô's formula on $A_t$ and viewing it as a function of $\mu_t$ and $p_t$, we obtain

$$\begin{aligned} dA_t = &\int (x-\mu_t)(x-\mu_t)^\top dp_t(x)dx - \int d\mu_t (x-\mu_t)^\top p_t(x)dx \\ &- \int (x-\mu_t)(d\mu_t)^\top p_t(x)dx \\ &- \frac{1}{2}\cdot 2\int (x-\mu_t)\, d\left[\mu_t^\top, p_t(x)\right]_t dx - \frac{1}{2}\cdot 2\int d\left[\mu_t, p_t(x)\right]_t (x-\mu_t)^\top dx \\ &+ \frac{1}{2}\cdot 2d\left[\mu_t, \mu_t^\top\right]_t \int p_t(x)dx. \end{aligned}$$

We observe that $\int d\mu_t (x-\mu_t)^\top p_t(x)dx = 0$ and $\int (x-\mu_t)(d\mu_t)^\top p_t(x)dx = 0$. Then,

$$\begin{aligned} d\left[\mu_t^\top, p_t(x)\right]_t &= (x-\mu_t)^\top A^{-1} A_t p_t(x)dt, \\ d\left[\mu_t, p_t(x)\right]_t &= A_t A^{-1}(x-\mu_t)\, p_t(x)dt, \\ d\left[\mu_t, \mu_t\right]_t &= A_t A^{-1} A_t dt. \end{aligned}$$

Combining all the terms together, we have

$$dA_t = \int (x-\mu_t)(x-\mu_t)^\top \left((x-\mu_t)^\top A^{-1/2} dW_t\right) p_t(x)dx - A_t A^{-1} A_t dt.$$

Finally, we derive the derivative of $\Gamma_t$. Define the function $\Gamma : \mathbb{R}^{d\times d} \mapsto \mathbb{R}$ as $\Gamma(X) = \operatorname{Tr}(X^q)$. The first-order and second-order derivatives of $\Gamma$ are given by

$$\left.\frac{\partial \Gamma}{\partial X}\right|_H = q\operatorname{Tr}\left(X^{q-1}H\right), \quad \left.\frac{\partial^2 \Gamma}{\partial X \partial X}\right|_{H_1,H_2} = q\sum_{a=0}^{q-2} \operatorname{Tr}\left(X^a H_2 X^{q-2-a} H_1\right).$$

Using the above derivatives and Itô's formula, we obtain

$$d\Gamma_t = d\operatorname{Tr}(Q_t^q) = q\operatorname{Tr}\left(Q_t^{q-1} dQ_t\right) + \frac{q}{2}\sum_{a=0}^{q-2}\sum_{i,j,k,l=1}^{d} \operatorname{Tr}\left(Q_t^a E_{ij} Q_t^{q-2-a} E_{kl}\right) d\left[Q_{ij}, Q_{kl}\right]_t, \tag{35}$$

where $E_{ij}$ is the matrix that takes 1 at the entry $(i,j)$ and 0 otherwise and $Q_{ij,t}$ is the stochastic process defined by the $(i,j)$ entry of $Q_t$. Using the derivative of $A_t$ in Equation (15), we have

$$\begin{aligned} dQ_t &= \int A^{-1/2}\left(x-\mu_t\right)\left(x-\mu_t\right)^\top A^{-1/2}\left(\left(x-\mu_t\right)^\top A^{-1/2} dW_t\right) p_t(x)dx \\ &\quad - Q_t^2 dt, \\ d\left[Q_{ij}, Q_{kl}\right]_t &= \int\int z(x)_i z(x)_j z(y)_k z(y)_l (x-\mu_t)^\top A^{-1}(y-\mu_t) p_t(x) p_t(y) dx dy dt, \end{aligned}$$

where $z(x)_i$ is the $i$th coordinate of $\left[A^{-1/2}(x-\mu_t)\right]$. Plugging the expressions of $dA_t$ and $d\left[A_{ij}, A_{kl}\right]_t$ into Equation (35), we obtain

$$\begin{aligned} d\Gamma_t &= q\int \left(x-\mu_t\right)^\top A^{-1/2}\left(Q_t\right)^{q-1} A^{-1/2}\left(x-\mu_t\right)\left(x-\mu_t\right)^\top A^{-1/2} dW_t p_t(x)dx \\ &\quad - q\operatorname{Tr}\left(Q_t^{q+1}\right)dt + \frac{q}{2}\sum_{a=0}^{q-2}\int\int \left(x-\mu_t\right)^\top A^{-1/2} Q_t^a A^{-1/2}\left(y-\mu_t\right) \\ &\quad \cdot \left(x-\mu_t\right)^\top A^{-1/2} Q_t^{q-2-a} A^{-1/2}\left(y-\mu_t\right)\left(x-\mu_t\right)^\top A^{-1}\left(y-\mu_t\right) p_t(x) p_t(y) dx dy dt. \end{aligned}$$

□

## References


[1] Z. Allen-Zhu, Y.T. Lee, and L. Orecchia. Using optimization to obtain a width-independent, parallel, simpler, and faster positive SDP solver. In: *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM (2016), pp. 1824–1831.

[2] M. Anttila, K. Ball, and I. Perissinaki. The central limit problem for convex bodies. *Transactions of the American Mathematical Society*, (12)355 (2003), 4723–4735

[3] K. Ball. Logarithmically concave functions and sections of convex sets in Rn. *Studia Math*, (1)88 (1988), 69–84

[4] J. Bourgain. On high dimensional maximal functions associated to convex bodies. *American Journal of Mathematics*, (6)108 (1986), 1467–1476

[5] H.J. Brascamp and E.H. Lieb. On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. In: *Inequalities*. Springer (2002), pp. 441–464.

[6] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In: *Proceedings of the Princeton Conference in Honor of Professor S. Bochner* (1969), pp. 195–199.

[7] B. Cousins and S. Vempala. A cubic algorithm for computing Gaussian volume. In: *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics (2014), pp. 1215–1228.

[8] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, (2)23 (2013), 532–569

[9] R. Eldan and B. Klartag. Approximately Gaussian marginals and the hyperplane conjecture. *Concentration, Functional Inequalities and Isoperimetry*, 545 (2011), 55–68

[10] R. Eldan and J. Lehec. Bounding the norm of a log-concave vector via thin-shell estimates. In: *Geometric Aspects of Functional Analysis*. Springer (2014), pp. 107–122.

[11] O. Guédon, P. Nayar, and T. Tkocz. Concentration inequalities and geometry of convex bodies. *Analytical and Probabilistic Methods in the Geometry of Convex Bodies*, 2 (2014), 9–86

[12] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, (3–4)13 (1995), 541–559

[13] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geometric & Functional Analysis GAFA*, (6)16 (2006), 1274–1290

[14] B. Klartag and V. Milman. The slicing problem by bourgain. In: *(To Appear) Analysis at Large, A Collection of Articles in Memory of Jean Bourgain*. Springer (2021).

[15] R. Latała and J. Wojtaszczyk. On the infimum convolution inequality. *Studia Mathematica*, (189)2 (2008), 147–187

[16] M. Ledoux. *The Concentration of Measure Phenomenon*, Number 89. American Mathematical Society (2001).

[17] Y.T. Lee and S.S. Vempala. Eldan's stochastic localization and the KLS hyperplane conjecture: an improved lower bound for expansion. In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE (2017), pp. 998–1007.

[18] Y.T. Lee and S.S. Vempala. The Kannan–Lovász–Simonovits conjecture. arXiv preprint arXiv:1807.03465 (2018).

[19] V.G. Maz'ya. Classes of domains and imbedding theorems for function spaces. In: *Doklady Akademii Nauk*, Vol. 133. Russian Academy of Sciences (1960), pp. 527–530.

[20] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Inventiones Mathematicae*, (1)177 (2009), 1–43

[21] B. Øksendal. *Stochastic Differential Equations*. Springer, Berlin (2003).

[22] G. Paouris. Concentration of mass on convex bodies. *Geometric & Functional Analysis GAFA*, (5)16 (2006), 1021–1049

[23] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, Vol. 293. Springer, Berlin (2013).

[24] A. Saumard and J.A. Wellner. Log-concavity and strong log-concavity: a review. *Statistics Surveys*, 8 (2014), 45

[25] P. Sternberg and K. Zumbrun. On the connectivity of boundaries of sets minimizing perimeter subject to a volume constraint. *Communications in Analysis and Geometry*, (1)7 (1999), 199–220

Y. Chen
Seminar for Statistics, ETH, Zurich, Switzerland.
yuansi.chen@stat.math.ethz.ch

Received: December 6, 2020
Revised: December 24, 2020
Accepted: December 27, 2020