Stochastic Localization and its Applications

Essay nr. 24. The Bourgain Slicing Problem

May 3, 2023

Contents

1 Introduction

The notion of stochastic localization was first introduced in the 2013 paper by Eldan [Eld13] in order to make progress regarding an isoperimetric problem known as the *Kannan-Lovász-Simonovitz* (KLS) conjecture. As it turns out, stochastic localization has been also useful in many other adjacent areas, in particular, in sampling and Markov mixing. This essay will provide an introduction to stochastic localization and describe its applications in Markov mixing and the KLS conjecture. Furthermore, using stochastic localization, this essay will also provide an alternative proof of a known bound for the log-Sobolev constant of log-concave measures.

Stochastic localization in its most general form describes a sequence of random measures (random variables taking values in the space of measures) which begins at a given measure and converges to dirac measures almost everywhere, namely it "localizes", and moreover, satisfy a certain martingale condition. These sequences of random measures are useful in studying specific measures. Namely, by evolving the stochastic localization in time, particular properties of the structure collapses allowing us say something about them. On the other hand, the martingale property allows us to preserve these properties (possibly up to some time). Hence, by balancing the two, i.e. allowing the sequence to evolve so it is close to a dirac measure while not evolve too long such that we lose the properties of the original measure, we are able to obtain results about the original measure.

In our case, we will mostly focus on one specific type of stochastic localization known as the linear-tilt localization. The linear-tilt localization is a special case of stochastic localizations in which at each time step, the measure is "tilted" in a random direction. This random direction can be chosen in a variety of ways however one choice of interest is when the random direction is chosen according to a Wiener process. A stochastic localization constructed this way accumulates a Gaussian component which becomes more and more significant as the process evolves. This is particularly helpful as Gaussian measures are well understood and we can use properties of the Gaussian to obtain results about our original measure.

1.1 Structure of this essay

This essay consists of an introduction to the theory of stochastic localization and subsequently presents three of it applications. We will now give a brief overview of these applications.

1.1.1 Markov mixing bounds

The first application of stochastic localization we will discuss is its application to bounding Markov mixing times. The motivation for Markov mixing bounds fundamentally comes from sampling. Suppose we wish to sample from some probability distribution μ. A common method to achieve this to through the use of the Markov chain Monte Carlo (MCMC):

Theorem 1 (MCMC)**.** Given (X_n) an irreducible positively recurrent homogenous Markov process on $\mathscr{X}$ with stationary distribution μ, for any $\phi : \mathscr{X} \to \mathbb{R}$ integrable,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \phi(X_k) = \int \phi \, \mathrm{d}\mu$$

almost everywhere.

With this theorem in mind, MCMC allows us to sample μ by sampling from a Markov chain instead. It is in general not difficult to come up with such Markov processes, although the difficulty often arises when we want to know its rate of convergence. This motivates the notion of mixing bounds which quantifies the time for which the Markov process takes before its law is approximately stationary.

Definition 1.1 (Total variation mixing time)**.** Given a probability measure $\nu \in \mathscr{M}(\mathscr{X})$, a Markov kernel K with stationary distribution μ and some $\epsilon > 0$, the ϵ-total variation mixing time is defined as

$$t_{\text{mix}}(P, \epsilon, \nu) := \inf\{t \geq 0 \mid \|K^t \nu - \mu\|_{\text{TV}} < \epsilon\},$$

Furthermore, we denote

$$t_{\text{mix}}(P, \epsilon) = \sup_{x \in \mathscr{X}} t_{\text{mix}}(P, \epsilon, \delta_x)$$

the worst mixing time starting at a point.

A standard method of analyzing the mixing times of Markov chains is through the use of the spectral gap.

Definition 1.2 (Spectral gap)**.** Given a Markov kernel K, we define its spectral gap to be

$$\mathfrak{gap}(K) := 1 - \sup\{\lambda \mid \lambda \text{ is an eigenvalue of } K, \lambda \neq 1\}.$$

Theorem 2 ([LPW17])**.** Given a reversible and irreducible Markov chain with kernel K on the state space $\mathscr{X}$ with stationary distribution μ, denoting $\mu_{\min} = \inf_{x \in \mathscr{X}} \mu(x)$, we have

$$t_{\text{mix}}(K, \epsilon) \leq \left\lceil \frac{1}{\mathfrak{gap}(K)} \left(\frac{1}{2} \log\left(\frac{1}{\mu_{\min}} \right) + \log\left(\frac{1}{2\epsilon} \right) \right) \right\rceil.$$

We will see that the spectral gap of Markov chains which kernels can be described using stochastic localizations is related to how the variance of the stochastic localization evolves. Thus, by analyzing the variance of the stochastic localization, in particular when the stochastic localization "conserves" variance, we can bound the spectral gap and consequently the mixing time of said kernel. As an example, we will apply this method to the Glauber dynamics on the Boolean hypercube to obtain a mixing time bound to the Ising model.

1.1.2 The KLS conjecture

We then move on to discuss a proof of Eldan's original 2013 result in [Eld13] which reduced the KLS conjecture to another seemingly weaker conjecture known as the thin-shell conjecture (up to an logarithmic factor). We quickly introduce these conjectures here.

Definition 1.3 (log-concave measure)**.** A measure μ on $\mathbb{R}^n$ is said to log-concave if it is of the form

$$\mathrm{d}\mu = e^{-V} \mathrm{dLeb}^n$$

for some convex function $V : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$.

Straightaway, we observe that the standard Gaussian measure on $\mathbb{R}^n$ is log-concave. Thus, as the Gaussian measures are very well understood, we are motivated to ask how similarly do log-concave measures behave when compared to the Gaussian. The KLS conjecture is one such comparison which compares the concentration of log-concave measures with that of the Gaussian.

Definition 1.4 (Concentration, [Eld18]). Let μ be a measure on $\mathbb{R}^n$, then μ is said to be C-concentrated if for all 1-Lipschitz function $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Var}_\mu[\phi] = \mathrm{Var}_{X\sim\mu}[\phi(X)] \leq \frac{1}{C^2}. \tag{1}$$

We denote the largest possible such C by C^μ_{con}.

As it turns out, the standard Gaussian measure on $\mathbb{R}^n$ is concentrated by a constant which is independent of the dimension n. Thus, we might make a conjecture of the form "all log-concave probability measures are concentrated by a universal constant". However, as currently stated, this statement is obviously false as spreading out the measure decreases the concentration. Furthermore, we need to be careful since, unlike the Gaussian, the concentration of general log-concave measures are not invariant under linear transformations. Hence, it is clear that the KLS conjecture would not hold without a suitable normalization. This leads us to the following.

Definition 1.5 (Isotropic). A measure μ on $\mathbb{R}^n$ is isotropic if $\mathbb{E}_{X\sim\nu}[X] = 0$ and $\mathrm{Cov}_{X\sim\nu}(X) = \mathrm{id}_n$.

Conjecture 1 (Kannan-Lovász-Simonovitz, [Eld18]). Denoting $\mathscr{M}^n_{\text{iso}}$ the set of all isotropic log-concave probability measures μ on $\mathbb{R}^n$ (recall that μ is isotropic if $\mathbb{E}_{X\sim\mu}[X] = 0$ and $\mathrm{Cov}_{X\sim\mu}(X) = \mathrm{id}_n$), there exists a *universal* constant C (i.e. does not depend on the dimension n) such that for all $\mu \in \mathscr{M}^n_{\text{con}}$, μ is C-concentrated.

From a more geometric point of view, the KLS conjecture can be equivalently phrased such that it asserts the maximum proportion of volume by surface area of a log-concave measure is bounded by a universal constant. As a result, the KLS conjecture has many important consequences in convex geometry. In particular, by noticing that any uniform measure on a convex body of unit volume is log-concave, the KLS conjecture directly implies the Bourgain slicing conjecture.

Conjecture 2 (Bourgain slicing, [Bou86]). For any convex body $U \subset \mathbb{R}^n$ of unit volume, there exists a hyperplane S such that $U \cap S$ has boundary measure of at least C for some universal constant C.

Moreover, as we shall see, the KLS conjecture in addition has applications to the Poincaré inequality, first moment concentration, exponential concentration and the aforementioned thin-shell conjecture. We will take a special look at the thin-shell conjecture which relaxes the KLS conjecture by only requiring the variance of the norm function to behave similar to the Gaussian.

Conjecture 3 (Thin-shell, [Eld13]). Taking $\mathscr{M}^n_{\text{iso}}$ as above, there exists a universal constant C such that for all $\mu \in \mathscr{M}^n_{\text{con}}$, we have

$$\sqrt{\mathrm{Var}_\mu[\|\cdot\|]} \leq \frac{1}{C}.$$

As a consequence of stochastic localization, we present a proof of the following theorem.

Theorem 3 (Eldan, [Eld13]). Denoting $\mathscr{M}^n_{\text{iso}}$ as above, we define

$$C^n_{\text{con}} := \sup\left\{C \mid \forall \mu \in \mathscr{M}^n_{\text{con}}, \mu \text{ is } C\text{-concentrated}\right\},$$

and

$$C^n_{\text{TS}} := \sup\left\{C \mid \forall \mu \in \mathscr{M}^n_{\text{con}}, \sqrt{\mathrm{Var}_\mu[\|\cdot\|]} \leq C^{-1}\right\},$$

we have,

$$C^n_{\text{con}} \leq C^n_{\text{TS}} \leq C^n_{\text{con}} \log n.$$

1.1.3 Log-Sobolev inequality

We will in the last section of this essay discuss the application of stochastic localization to the log-Sobolev inequality. Similar to the KLS conjecture, we are interested to compare the log-concave measures to the Gaussian measure. In this case, motivated by the Gaussian log-Sobolev inequality, we establish a similar inequality for log-concave measures bounding the entropy of said measures. Heuristically, similar to that of the variance, the entropy of a random variable is a measure of its uncertainty or randomness.

Definition 1.6 (Entropy)**.** Given $\phi : \mathscr{X} \to \mathbb{R}_{\geq 0}$ and a measure μ, we define the entropy of ϕ with respect to μ to be

$$\mathrm{Ent}_\mu[\phi] := \mathbb{E}_\mu\left[\phi \log\left(\frac{1}{\mathbb{E}_\mu[\phi]}\phi\right)\right] = \int \phi \log \phi \, \mathrm{d}\mu - \int \phi \, \mathrm{d}\mu \log\left(\int \phi \, \mathrm{d}\mu\right)$$

with the convention that $0 \log 0 = 0$.

The log-Sobolev inequality is then formulated as the following.

Definition 1.7 (Log-Sobolev inequality, [LV16])**.** For a given measure μ on $\mathbb{R}^n$, μ is said to satisfy the log-Sobolev inequality with log-Sobolev constant ρ_μ if ρ_μ is the largest ρ such that for all smooth $\phi : \mathbb{R}^n \to \mathbb{R}$ with $\int \phi^2 \mathrm{d}\mu = 1$, we have

$$\frac{\rho}{2}\mathrm{Ent}_\mu[\phi^2] \leq \mathbb{E}_\mu[\|\nabla \phi\|^2] = \int \|\nabla \phi\|^2 \mathrm{d}\mu.$$

The log-Sobolev inequality is an incredibly useful inequality while studying the concentration of measures. In particular, should a measure satisfy a log-Sobolev equality, by using a well-known method commonly known as Herbst's argument, one can obtain the exponential concentration of said measure via the Chernoff bound.

Theorem 4 (Herbst's argument)**.** If μ satisfy the log-Sobolev inequality with log-Sobolev constant ρ_μ, then for all ϕ with uniformly bounded gradient $\|\nabla \phi\| \leq K$, we have

$$\psi^\mu_{\phi - \mathbb{E}_\mu[\phi]}(\lambda) \leq \frac{K^2 \lambda^2}{2\rho_\mu}$$

where $\psi^\mu_{\phi - \mathbb{E}_\mu \phi}$ is the logarithmic moment generating function of $\phi - \mathbb{E}_\mu[\phi]$ with respect to μ.

A proof of this theorem is included in Appendix A.

It does not come as a surprise that as with the setting of the KLS conjecture, the standard Gaussian measure γ^n in $\mathbb{R}^n$ satisfies the log-Sobolev inequality with log-Sobolev constant $\rho_{\gamma^n} = 1$ (c.f. [Gro75]). However, in contrast to the KLS conjecture which suggests that the concentration of log-concave measures are bounded below by a universal constant, it is known (c.f. [LV16]) that the log-Sobolev constant cannot be bounded below by a universal constant. Nonetheless, we are interested in the log-Sobolev constant of log-concave measure. More specifically, we will study the log-Sobolev constant for log-concave measures supported in a ball of fixed diameter to obtain the following result.

Theorem 5 ([LV16])**.** For any isotropic log-concave measure μ on $\mathbb{R}^n$ with support on a ball of diameter D, μ has log-Sobolev constant $\rho_\mu \gtrsim D^{-1}$.

2 Stochastic Localization Scheme

In this section we introduce the notion of stochastic localization and provide some useful examples of them which are studied further in the subsequent sections.

We work in general Borel spaces $(\mathscr{X}, \Sigma)$ for this section while restricting our focus to either the Euclidean space $\mathbb{R}^n$ or the Boolean hypercube $\{-1, 1\}^n$ in subsequent sections. We take $(\Omega, \mathscr{F}, \mathbb{P})$ to be our underlying probability space and we introduce the notation $\mathscr{M}(\mathscr{X})$ for the space of probability measures on $\mathscr{X}$.

Definition 2.1 (Prelocalization process)**.** Given $\mu \in \mathscr{M}(\mathscr{X})$, a measure-valued stochastic process $(\mu_t)_{t\geq 0}$ is said to be a prelocalization of μ if

- L0 $\mu_0 = \mu$.
- L1 For all $t \geq 0$, μ_t is a probability measure almost everywhere, i.e. $\mathbb{P}(\mu_t(\mathscr{X}) = 1) = 1$.
- L2 For all $A \in \Sigma$, $(\mu_t(A))_{t\geq 0}$ is a martingale with respect to the natural filtration of (μ_t).

Definition 2.2 (Stochastic localization process, [CE22])**.** Given $\mu \in \mathscr{M}(\mathscr{X})$, a measure-valued stochastic process $(\mu_t)_{t\geq 0}$ is said to be a stochastic localization of μ if in addition to being a prelocalization of μ, $(\mu_t)_{t\geq 0}$ also satisfies

- L3 For all $A \in \Sigma$, $\mu_t(A)$ converges almost everywhere to 0 or 1 as $t \to \infty$.

We say a stochastic localization is discrete if t takes value in $\mathbb{N}$ and continuous if t takes value in $\mathbb{R}_{\geq 0}$. For shorthand, we denote $(\mu_k)_k$ for a discrete stochastic localization of μ.

Proposition 2.1. Straightaway, by the martingale property, if $(\mu_t)_{t\geq 0}$ is a stochastic localization of μ, then

- $\mathbb{E}[\mu_t] = \mu$ for all $t \geq 0$.
- $\mu_t \to \delta_X$ almost everywhere as $t \to \infty$ for some $X \sim \mu$ (here weak and total variational convergence are equivalent and so $\to$ can mean either).

Proof. The first statement is immediate as for all $A \in \Sigma$,

$$\mathbb{E}[\mu_t](A) \triangleq \mathbb{E}[\mu_t(A)] = \mathbb{E}[\mu_0(A)] = \mu(A).$$

To prove the second statement, let us first parse what the claim is. Fixing a realization ω of μ_t, we have by (L3) that μ_t converges to some Dirac measure based at some $x_\omega \in \mathscr{X}$. Thus, defining the random variable $X : \omega \mapsto x_\omega$, we need to show $X \sim \mu$. Indeed, by taking $\phi : \mathscr{X} \to \mathbb{R}$ to be any bounded and continuous function, by the definition of X

$$\int \phi(x)\mu_t(\mathrm{d}x) \xrightarrow{\text{a.e.}} \int \phi(x)\delta_X(\mathrm{d}x) = \phi(X) \text{ as } t \to \infty.$$

Thus, taking expectation on both sides, we have

$$\mathbb{E}[\phi(X)] = \mathbb{E}\left[\int \phi \,\mathrm{d}\mu_t\right] = \int \phi \,\mathrm{d}\mathbb{E}[\mu_t] = \int \phi \,\mathrm{d}\mu$$

implying $X \sim \mu$ as required. $\square$

An example of a stochastic localization scheme is the coordinate by coordinate localization scheme on $\mathscr{X} = \{-1,1\}^n$. This scheme relates to the Glauber dynamics for which the stochastic localization scheme provides a mixing bound. We will discuss this property further in Section 3.3 while only considering its construction for now.

Given a probability measure μ on $\{-1,1\}^n$, we introduce the random variable $X \sim \mu$, and Y a uniform random variable over all permutations of $[n] = \{1,\dots,n\}$ independent of X. Then, the coordinate by coordinate stochastic localization of μ is the process $(\mu_k)_k$ such that for all $x \in \{-1,1\}^n$,

$$\mu_k(x) = \mathbb{P}(X = x \mid X_{Y_1},\dots,X_{Y_{n\wedge k}}).$$

Namely, μ_k is the law of X conditioned on $X_{Y_1},\dots,X_{Y_i}$.

$(\mu_k)_k$ is indeed a stochastic localization of μ. It is clear that (L0) and (L1) are satisfied. By construction of $(\mu_k)_k$, denoting $\mathscr{F}_k := \sigma(X_{Y_1},\dots,X_{Y_{n\wedge k}})$, we have by the tower property

$$\mathbb{E}[\mu_{k+1}(x) \mid \mathscr{F}_k] = \mathbb{E}[\mathbb{E}[\mathbb{P}(X = x \mid X) \mid \mathscr{F}_{k+1}] \mid \mathscr{F}_k] = \mathbb{E}[\mathbb{P}(X = x \mid X) \mid \mathscr{F}_k] = \mu_k(x)$$

implying $(\mu_k(x))$ a martingale as required for (L2). Finally, it is clear that

$$\lim_{k\to\infty} \mu_k(x) = \mu_n(x) = \mathbb{P}(X = x \mid X) = \mathbf{1}_{\{X=x\}} \in \{0,1\}$$

implying (L3).

An analogous construction of the coordinate by coordinate stochastic localization scheme in $\mathbb{R}^n$ is the random subspace localization. Similar to before, for a probability measure μ on $\mathbb{R}^n$, we introduce the random variable $X \sim \mu$ and Y a uniform random variable on $O(n)$ (so the column vectors $\{Y_1,\dots,Y_n\}$ form an orthonormal basis of $\mathbb{R}^n$) independent of X. Then, we define the random subspace stochastic localization of μ as $(\mu_k)_k$ where μ_k is the law of X conditioned on $\langle X, Y_1\rangle,\dots,\langle X, Y_{n\wedge k}\rangle$.

2.1 Linear-tilt localization schemes

An important class of stochastic localization schemes are the linear-tilt schemes. Introduced by Eldan in [Eld13], linear-tilt schemes has been vital in the recent progress regarding the KLS conjecture. More recently, a discrete version of the linear-tilt scheme was introduced in [CE22] and is used to provide a mixing bound for Glauber dynamics. We will in this section introduce this family of localizations and consider two specific examples of such linear-tilt schemes which are useful for our analysis later.

Informally, given a probability measure μ on $\mathscr{X} \subseteq \mathbb{R}^n$, the linear-tilt scheme of μ is constructed recursively in which at each step, we pick a random direction and multiply the density at this time with a linear function along this direction (i.e. a tilt along a random direction).

Let μ be a probability measure on $\mathscr{X} \subseteq \mathbb{R}^n$, we introduce the following definition.

Definition 2.3 (Barycenter)**.** The barycenter of μ with respect to the function $F : \mathbb{R}^n \to \mathbb{R}$ is

$$\bar{\mu}(F) := \int_{\mathscr{X}} xF(x)\mu(\mathrm{d}x).$$

In the case that $F = \mathrm{id}$, we simply write $\bar{\mu} = \bar{\mu}(F) = \mathbb{E}_{X\sim\mu}[X]$.

Definition 2.4 (Linear-tilt localization)**.** A measure-valued stochastic process $(\mu_t)_{t\geq 0}$ is said to be a linear-tilt localization of the probability measure μ if

1. $\mu_t \ll \mu$ for each $t \geq 0$, and
2. denoting $F_t := \mathrm{d}\mu_t/\mathrm{d}\mu$, we have $F_0 = 1$ and

$$\mathrm{d}F_t(x) = \langle x - \bar{\mu}(F_t), \mathrm{d}Z_t\rangle F_t(x) \tag{2}$$

for some stochastic process $(Z_t)_{t\geq 0}$ such that $\mathbb{E}[\mathrm{d}Z_t \mid \mu_t] = 0$ for all $t \geq 0$.

It is clear that $(\mu_t(A))_t$ is a martingale for all $A \in \Sigma$ by observing that Equation (2) has no drift term. Furthermore, since $\mu_t(\mathscr{X}) = \int F_t \mathrm{d}\mu$ is differentiable by construction, $(\mu_t(\mathscr{X}))$ has zero quadratic variation and thus is constant in t. With this in mind, as $\mu_0 = \mu$ is a probability measure, it follows:

Proposition 2.2. If $(\mu_t)_t$ is a linear-tilt localization of μ, then μ_t is a probability measure for each t.

Corollary 2.3. A linear-tilt localization $(\mu_t)_t$ of μ is a prelocalization of μ.

We remark that in general, a linear-tilt localization is not necessarily a stochastic localization as (L3) might not be satisfied. It is possible to impose sufficient conditions on (Z_t) for which (L3) holds, e.g. by requiring $\|\mathrm{Cov}(Z_t)\|_{\mathrm{op}}$ to decrease sufficiently fast. However, for generality, we will not restrict ourselves to one of these conditions. Instead, we will consider (L3) case by case in the following examples of linear-tilt schemes.

2.1.1 Linear-tilt localization driven by a Wiener process

A natural choice of $(Z_t)_{t\geq 0}$ is the standard Wiener process on $\mathbb{R}^n$. Denoting $(W_t)_{t\geq 0}$ a standard Wiener process on $\mathbb{R}^n$, we define the random functions $(F_t)_{t\geq 0}$ to be the solution of the following infinite system of SDEs (existence and uniqueness is established by theorem 5.2 in [Øks03]):

$$F_0 = 1, \mathrm{d}F_t(x) = \langle x - \bar{\mu}(F_t), \mathrm{d}W_t\rangle F_t(x), \tag{3}$$

for all $x \in \mathbb{R}^n$. We shall from this point forward denote the random variables $a_t := \bar{\mu}(F_t)$.

By applying Itô's formula, we make the following useful observation: for all $x \in \mathbb{R}^n$,

$$\mathrm{d}\log F_t(x) = \frac{\mathrm{d}F_t(x)}{F_t(x)} - \frac{\mathrm{d}[F(x)]_t}{2F_t(x)^2} = \langle x - a_t, \mathrm{d}W_t\rangle - \frac{1}{2}\|x - a_t\|^2 \mathrm{d}t \tag{4}$$

where the second equality follows by the construction of F. Hence, as $\log F_0(x) = 0$, we observe

$$\begin{aligned}
\log F_t(x) &= \int_0^t \langle x - a_s, \mathrm{d}W_s\rangle - \frac{1}{2}\int_0^t \|x - a_s\|^2 \mathrm{d}s \\
&= \left(\langle x, W_t\rangle - \int_0^t \langle a_s, \mathrm{d}W_s\rangle\right) - \left(\frac{t}{2}\|x\|^2 + \frac{1}{2}\int_0^t \|a_s\|^2 \mathrm{d}s - \int_0^t \langle x, a_s\rangle \mathrm{d}s\right) \\
&= -\left(\int_0^t \langle a_s, \mathrm{d}W_s\rangle + \frac{1}{2}\|a_s\|^2 \mathrm{d}s\right) + \langle x, a_t + W_t\rangle - \frac{t}{2}\|x\|^2.
\end{aligned}$$

Thus, taking $\mathrm{d}z_t := \langle a_t, \mathrm{d}W_t\rangle + \frac{1}{2}\|a_t\|^2 \mathrm{d}t$ and $v_t := a_t + W_t$, we observe $F_t(x)$ is a density function with a Gaussian component of the form

$$F_t(x) = e^{z_t + \langle x, v_t\rangle - \frac{t}{2}\|x\|^2}, \tag{5}$$

for given Itô processes $(z_t), (v_t)$.

With this formulation of $F_t(x)$ in mind, it follows that F_t is non-negative, and so, we may define $(\mu_t)_t$ to be the process such that $\mathrm{d}\mu_t = F_t \mathrm{d}\mu$. It is clear that $(\mu_t)_t$ is a linear-tilt localization of μ and so, is a prelocalization of μ.

The remainder of this section is devoted to showing $(\mu_t)_t$ is furthermore a stochastic localization of μ in the special case that it is log-concave (namely we will show (L3) for this special case), and prove some basic properties about this process useful for our analysis later.

Straightaway, we observe that if μ is log-concave, μ_t has density of the form $F_t e^{-V}$ with respect to the Lebesgue measure. Thus, as t increases, we observe that μ_t becomes gradually more akin to a Gaussian measure. This fact is very beneficial for our analysis later.

To show (μ_t) satisfies (L3) if μ is log-concave, we study the limiting behavior of (μ_t) as $t \to \infty$ by considering their covariances:

$$A_t := \mathrm{Cov}[\mu_t] = \int (x - a_t) \otimes (x - a_t) \mu_t(\mathrm{d}x), \tag{6}$$

where $\otimes$ denotes the Kronecker product. In particular, we will show $(A_t)_{ij} \to 0$ for all $i, j \in \{1, \ldots, n\}$ as $t \to \infty$ allowing us to conclude (μ_t) converges weakly to some Dirac measure. Indeed, this is a direct consequence of the following lemma.

Lemma 2.4 (Brascamp-Lieb, [BL76])**.** Given $V : \mathbb{R}^n \to \mathbb{R}$ convex and $K > 0$, if ν is an isotropic probability measure on $\mathbb{R}^n$ of the form

$$\mathrm{d}\nu = Z e^{-V(x) - \frac{K}{2}\|x\|^2} \mathrm{dLeb}^n(x)$$

with Z being the normalization constant, then ν satisfy the following (Poincaré) inequality: for all smooth ϕ,

$$\mathrm{Var}_\nu[\phi] \le K^{-1} \mathbb{E}_\nu[\|\nabla \phi\|^2].$$

With this lemma in mind, by taking $\nu = \mu_t$ using Equation (5) and defining $\pi_i(x) := x_i$, we have by the Cauchy-Schwarz inequality

$$(A_t)_{ij} \le \sqrt{\mathrm{Var}_{\mu_t}[\pi_i]} \sqrt{\mathrm{Var}_{\mu_t}[\pi_j]} \le \max_{k=1,\ldots,n} \frac{1}{t} \int \|\nabla \pi_k\|^2 \mathrm{d}\mu_t$$

Again, using Equation (5), we note that any realizations of $(F_t(x))$ is eventually decreasing in t for all $x \neq 0$, implying

$$\sup_{t>0} \max_{k=1,\ldots,n} \int \|\nabla \pi_k\|^2 \mathrm{d}\mu_t = \sup_{t>0} \max_{k=1,\ldots,n} \int x_k^2 \mathrm{d}\mu_t < \infty.$$

Thus, by taking $t \to \infty$ we have $(A_t)_{ij} \to 0$ for all $i, j \in \{1, \ldots, n\}$ as claimed and we have (μ_t) satisfying (L3). Hence, the linear-tile localization is indeed a stochastic localization of μ.

Corollary 2.5. (μ_t) converges set-wise to some Dirac measure almost everywhere. We denote this limiting (random) Dirac measure by δ_{a_∞} where a_∞ is some $\mathbb{R}^n$-valued random variable.

As a result of Proposition 2.1, we also have the following corollaries.

Corollary 2.6. The massive point a_∞ of the limiting Dirac measure is the limit of a_t as $t \to \infty$ and has law μ.

Corollary 2.7. Taking ϕ to be any continuous function, we define $M_t = \int \phi \mathrm{d}\mu_t$. Then, $(M_t)_{t\geq 0}$ is a martingale and

$$M_t \xrightarrow{\text{a.e.}} \phi(a_\infty) \sim \phi_*\mu \tag{7}$$

where $\phi_*\mu$ denotes the push-forward measure of μ along ϕ.

2.1.2 Discrete time linear-tilt localization

We may construct an analogous version of the linear-tilt localization for discrete time. By utilizing the little-o notation, Equation (2) can be rewritten as

$$\frac{\mathrm{d}\mu_{t+h}}{\mathrm{d}\mu}(x) = \frac{\mathrm{d}\mu_t}{\mathrm{d}\mu}(x) + \langle x - \bar{\mu}(F_t), h\mathrm{d}Z_t\rangle \frac{\mathrm{d}\mu_t}{\mathrm{d}\mu}(x) + o(h).$$

Hence, an discrete analog of the linear tilt localization is defined as the following.

Definition 2.5 (Discrete time linear-tilt localization)**.** Given a measure $\mu \in \mathscr{M}(\mathscr{X})$, the discrete time linear-tilt localization is the sequence of random measures $(\mu_k)_k$ defined by $\mu_0 = \mu$ and

$$\mathrm{d}\mu_{k+1} = (1 + \langle x - \bar{\mu}_k, Z_k\rangle)\mathrm{d}\mu_k \tag{8}$$

for some sequence of random variables such that $\mathbb{E}[Z_k \mid \mu_k] = 0$ for all $k \in \mathbb{N}$.

The coordinate by coordinate localization can be formulated as a discrete time linear-tilt localization. Given μ a probability measure on $\{-1,1\}^n$, we recall that the coordinate by coordinate localization is defined by "pinning" an additional random coordinate after each time step. To phrase this as a linear-tilt localization, we take the random variables Z_k to be

$$Z_k := e_{Y_k} \cdot \begin{cases} \frac{1}{1+(\bar{\mu}_k)_{Y_k}} & \text{with probability } \frac{1+(\bar{\mu}_k)_{Y_k}}{2} \\ \frac{-1}{1-(\bar{\mu}_k)_{Y_k}} & \text{with probability } \frac{1-(\bar{\mu}_k)_{Y_k}}{2} \end{cases} \tag{9}$$

where $e_1, \dots, e_n$ are the standard basis of $\mathbb{R}^n$ and again Y is a uniform random variable over all permutations of $[n]$. Thus, the linear-tilt localization (μ_k) given by this choice of (Z_k) is defined by

$$\mu_{k+1}(\sigma) = (1 + \|Z_k\|(\sigma_{Y_k} - (\overline{\mu}_k)_{Y_k}))\mu_k(\sigma)$$

for all $k < n$. Similar to before, we terminate the process at time n and so we extend the process to all times by taking $\mu_k = \mu_{n\wedge k}$.

Let us parse this definition to see why this is equivalent to the coordinate by coordinate localization. Taking $k < n$, we have at the $k+1$-th step, Y_{k+1} chooses a random axis which had not been chosen before. Then, Z_k is chosen such that for each configuration σ, the probability of σ_{Y_k} being ± 1 is proportional the mass of μ_k on $\pm e_{Y_k}$. This is precisely the steps needed to construct the coordinate by coordinate localization as conditioning on an additional axis in this case is simply a projection on to said axis.

3 Markov Mixing

An application of stochastic localizations is used to prove mixing bounds for Markov processes. Focusing on the framework established by Chen and Eldan in [CE22], we will present their prescribed method to bound the mixing time for a special class of Markov processes: the Markov chains associated with stochastic localizations. As a specific example, we will in this section also describe Chen and Eldan's [CE22] application of this framework to the Glauber dynamics and as a result, recover a version of the main theorem on spectral independence as presented by Anari, Liu and Oveis Gharan in [ALOG20].

3.1 Ising model and Glauber dynamics

Before moving directly to establishing the framework, let us first quickly discuss the Ising model and the main result of this section. The Ising model is a well-known measure in mathematical physics and is often used to model magnetism and gas behavior. While fundamentally, the Ising model is constructed using Bernoulli random variables, as these random variables interacts with one another, sampling the Ising model directly has exponential computational complexity increasing with the number of particles. Hence, to sample from the Ising model, one often instead applies MCMC to the Glauber dynamics which is a Markov chain with stationary measure being the Ising model. We will in this small section introduce these definitions and recall a known bound from [ALOG20] for the spectral gap and consequently the mixing time of the Glauber dynamics.

Heuristically, the Ising model measures the probability that a graph is in a specific configuration of up and down spins in which neighboring vertices are more likely to have the same spin. This "likeliness" is controlled by a parameter β in which a larger β means that neighboring vertices are more likely to align.

Definition 3.1 (Ising model, [CE22]). Given a graph $G = (V, E)$, $\beta > 0$ and $h \in \mathbb{R}$, the Ising model on G with inverse temperature β and external field h is the probability measure $\mu_{\beta,h}$ on $\{-1,1\}^V$ defined such that for all $\sigma \in \{-1,1\}^V$,

$$\mu_{\beta,h}(\sigma) := \frac{1}{Z} \exp\left\{\beta \sum_{xy \in E} \sigma_x \sigma_y + h \sum_{x \in V} \sigma_x\right\}$$

where $Z > 0$ is the normalizing constant.

To sample from the Ising model, one method is by starting from an arbitrary configuration $\sigma_1 \in \{-1,1\}^n$, and at each time step, update the configuration according to the description of the Ising model. Namely, the configuration at the next time step either remains the same or change at one vertex. Furthermore, the probability of this occurring is weighted according to a given measure μ. This procedure results in a Markov chain which turns out to have the Ising model as its stationary distribution. This Markov chain is known as the Glauber dynamics. Formally, it is defined as the following.

Definition 3.2 (Glauber dynamics, [CE22]). Given a measure μ of $\{-1,1\}^n$, the Glauber dynamics of μ is the Markov chain with kernel

$$K(\sigma^1, \sigma^2) := \mathbf{1}_{\{\|\sigma^1-\sigma^2\|_1=1\}} \frac{1}{n} \frac{\mu(\sigma^2)}{\mu(\sigma^1)+\mu(\sigma^2)} + \mathbf{1}_{\{\sigma^1=\sigma^2\}} \frac{1}{n} \sum_{\|\tilde{\sigma}-\sigma^1\|_1=1} \frac{\mu(\sigma^1)}{\mu(\sigma^1)+\mu(\tilde{\sigma})}, \tag{10}$$

where we define $\|\sigma^1-\sigma^2\|_1 := \frac{1}{2}\sum_{i=1}^n |\sigma_i^1-\sigma_i^2|$ for all $\sigma^1,\sigma^2 \in \{-1,1\}^n$.

Recently, Anari, Liu and Oveis Gharan in [ALOG20] were able to provide a bound for the spectral gap for a special class of Glauber dynamics satisfying a property known as spectral independence. We quickly recall their results here.

Definition 3.3 (Pairwise influence matrix, [ALOG20])**.** Given a measure μ on $\{-1,1\}^n$, we define the pairwise influence matrix $\Psi(\mu)$ such that it has entries

$$\Psi(\mu)_{i,j} := \mathbb{P}_{X\sim\mu}[X_i = 1 \mid X_j = 1] - \mathbb{P}_{X\sim\mu}[X_i = 1 \mid X_j = -1],$$

for all $i \neq j$ and we set $\Psi(\mu)_{i,i} = 0$ for all i.

Definition 3.4 (Spectral independence, [ALOG20])**.** A measure μ on $\{-1,1\}^n$ is said to be η-spectrally independent if the maximum eigenvalue of $\Psi(\mu)$ is bounded above by η.

In addition, we say μ is $(\eta_0,\ldots,\eta_{n-2})$-spectrally independent if defining μ_k to be the law of X conditioned on $X_{i_1},\ldots,X_{i_k}$ for some $X\sim\mu$ and any $\{i_1,\ldots,i_k\}\subseteq\{1,\ldots,n\}$, μ_k is η_k-spectrally independent for each $k=0,\ldots,n-2$.

Theorem 6 ([ALOG20])**.** If μ is a measure on $\{-1,1\}^n$ which is $(\eta_0,\ldots,\eta_{n-2})$-spectrally independent, then the Glauber dynamics of μ has spectral gap bound

$$\mathfrak{gap}(K) \geq \frac{1}{n}\prod_{k=0}^{n-2}\left(1-\frac{\eta_k}{n-k-1}\right).$$

As an application of stochastic localizations, we will in Section 3.3 recover a similar bound by using the theories established above. In particular, assuming the same assumptions as Theorem 6, we will show that

$$\mathfrak{gap}(K) \geq \prod_{k=0}^{n-2}\left(1-\frac{\eta_k+1}{n-k}\right).$$

3.2 Dynamics of stochastic localizations

As alluded to in the introduction, one may associate a Markov process at each time step of a stochastic localization process for which the original process is stationary. We will in this section define these Markov processes and show that the Glauber dynamics can be constructed using this method.

Definition 3.5 (Markov process associated with a stochastic localization, [CE22])**.** Let $(\mu_t)_{t\geq 0}$ be a prelocalization of μ such that μ_t is absolutely continuous with respect to μ for all t. Then, for all $\tau > 0$, we define the dynamics associated with $(\mu_t)_t$ at τ to be the Markov process with kernel

$$K(x,A) := \mathbb{E}_{\mathbb{P}}\left[\frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}(x)\mu_\tau(A)\right]$$

for all $x\in\mathcal{X}, A\in\Sigma$.

As hinted by the notation, rather than a deterministic time τ, τ can also be taken to be an appropriate stopping time. In this case, the theorems below will remain to hold by invoking the optional stopping theorem whenever necessary.

This construction indeed results in a kernel since for each $x \in \mathscr{X}$, as $\mathbb{E}_{\mathbb{P}}[\mu_\tau] = \mu$ by (L2),

$$K(x,\Omega) = \mathbb{E}_{\mathbb{P}}\left[\frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}(x)\right] = \frac{\mathrm{d}}{\mathrm{d}\mu}\mathbb{E}_{\mathbb{P}}[\mu_\tau](x) = \frac{\mathrm{d}\mu}{\mathrm{d}\mu}(x) = 1$$

where the third equality follows by the uniqueness of the Radon-Nikodym derivative. In particular, for all $A \in \Sigma$, we have

$$\int_A \mathbb{E}_{\mathbb{P}}\left[\frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}\right]\mathrm{d}\mu = \mathbb{E}_{\mathbb{P}}\left[\int_A \frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}\mathrm{d}\mu\right] = \mathbb{E}_{\mathbb{P}}[\mu_\tau](A).$$

Proposition 3.1. The Markov process associated with a prelocalization $(\mu_t)_{t\geq 0}$ of μ is reversible and has stationary distribution μ.

Proof. Taking $\phi : \mathscr{X}^2 \to \mathbb{R}$ integrable, we have by Fubini's theorem

$$\begin{aligned}\int_{\mathscr{X}^2} \phi(x,y)K(x,\mathrm{d}y)\mu(\mathrm{d}x) &= \int \phi(x,y)\mathbb{E}_{\mathbb{P}}\left[\frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}(x)\mu_\tau(\mathrm{d}y)\right]\mu(\mathrm{d}x) \\ &= \mathbb{E}_{\mathbb{P}}\left[\iint \phi(x,y)\mu_\tau(\mathrm{d}y)\frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}(x)\mu(\mathrm{d}x)\right] \qquad (11) \\ &= \mathbb{E}_{\mathbb{P}}\left[\iint \phi(x,y)\mu_\tau(\mathrm{d}y)\mu_\tau(\mathrm{d}x)\right].\end{aligned}$$

Similarly, by the same calculation,

$$\int_{\mathscr{X}^2} \phi(x,y)K(y,\mathrm{d}x)\mu(\mathrm{d}y) = \mathbb{E}_{\mathbb{P}}\left[\iint \phi(x,y)\mu_\tau(\mathrm{d}y)\mu_\tau(\mathrm{d}x)\right].$$

Thus,

$$\int_{\mathscr{X}^2} \phi(x,y)K(x,\mathrm{d}y)\mu(\mathrm{d}x) = \int_{\mathscr{X}^2} \phi(x,y)K(y,\mathrm{d}x)\mu(\mathrm{d}y)$$

for any integrable $\phi : \mathscr{X}^2 \to \mathbb{R}$ implying K is reversible.
On the other hand, for all $A \in \Sigma$, we compute using the martingale property

$$\begin{aligned}K\mu(A) &= \int K(x,A)\mu(\mathrm{d}x) = \int \mathbb{E}_{\mathbb{P}}\left[\frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}(x)\mu_\tau(A)\right] = \mathbb{E}_{\mathbb{P}}\left[\mu_\tau(A)\int \frac{\mathrm{d}\mu_\tau}{\mathrm{d}\mu}\mathrm{d}\mu\right] \\ &= \mathbb{E}_{\mathbb{P}}[\mu_\tau(A)\mu_\tau(\Omega)] = \mathbb{E}_{\mathbb{P}}[\mu_\tau(A)] = \mu(A)\end{aligned}$$

implying μ is the stationary measure of K. □

Proposition 3.2. Taking K to be the kernel of the Markov process associated with a prelocalization $(\mu_t)_{t\geq 0}$ of μ at time τ, we have

$$\mathfrak{gap}(K) = \inf_{\phi:\mathscr{X}\to\mathbb{R}} \frac{\mathbb{E}[\mathrm{Var}_{\mu_\tau}[\phi]]}{\mathrm{Var}_\mu[\phi]}.$$

Proof. By Equation (11) (taking $\phi(x,y)=\phi(x)\phi(y)$), we have

$$\int_{\mathscr{X}^2} \phi(x)\phi(y)K(x,\mathrm{d}y)\mu(\mathrm{d}x) = \mathbb{E}_{\mathbb{P}}\left[\left(\int_{\mathscr{X}} \phi\,\mathrm{d}\mu_\tau\right)^2\right], \tag{12}$$

for any integrable $\phi : \mathscr{X} \to \mathbb{R}$. On the other hand,

$$\int_{\mathscr{X}^2} \phi(y)^2 K(x,\mathrm{d}y)\mu(\mathrm{d}x) = \int (K\phi^2)(x)\mu(\mathrm{d}x) = \int \phi(x)^2(K\mu)(\mathrm{d}x) = \int \phi(x)^2\mu(\mathrm{d}x)$$

as μ is the stationary measure of K. Thus, for any integrable $\phi : \mathscr{X} \to \mathbb{R}$, by substituting the above two equations, we have

$$\begin{aligned}
&\frac{1}{2\mathrm{Var}_\mu[\phi]}\int_{\mathscr{X}^2}(\phi(x)-\phi(y))^2K(x,\mathrm{d}y)\mu(\mathrm{d}x)\\
=&\frac{1}{2\mathrm{Var}_\mu[\phi]}\left(\int \phi(x)^2\mu(\mathrm{d}x) - 2\int \phi(x)\phi(y)K(x,\mathrm{d}y)\mu(\mathrm{d}x) + \int \phi(y)^2K(x,\mathrm{d}y)\mu(\mathrm{d}x)\right)\\
=&\frac{1}{\mathrm{Var}_\mu[\phi]}\left(\int \phi^2\mathrm{d}\mu - \mathbb{E}_{\mathbb{P}}\left[\left(\int_{\mathscr{X}}\phi\,\mathrm{d}\mu_\tau\right)^2\right]\right)\\
=&\frac{1}{\mathrm{Var}_\mu[\phi]}\mathbb{E}_{\mathbb{P}}\left[\int \phi^2\mathrm{d}\mu_\tau - \left(\int_{\mathscr{X}}\phi\,\mathrm{d}\mu_\tau\right)^2\right] = \frac{\mathbb{E}[\mathrm{Var}_{\mu_\tau}[\phi]]}{\mathrm{Var}_\mu[\phi]}.
\end{aligned} \tag{13}$$

Moreover, by observing that

$$\mathfrak{gap}(K) = \inf_{\substack{\phi:\mathscr{X}\to\mathbb{R}\\ \int \phi\,\mathrm{d}\mu=0}} 1 - \frac{\int \phi K\phi\,\mathrm{d}\mu}{\int \phi^2\mathrm{d}\mu} = \inf_{\phi:\mathscr{X}\to\mathbb{R}} \frac{1}{2\mathrm{Var}_\mu[\phi]}\int (\phi(x)-\phi(y))^2K(x,\mathrm{d}y)\mu(\mathrm{d}x),$$

the result follows by taking infimum on both sides of (13). ☐

With regards to Theorem 2, in the case μ has full support on the finite state space $\mathscr{X}$ (e.g. in the setting of Glauber dynamics) the above proposition provides a method for computing an upper bound for the mixing time. In particular, should the prelocalization $(\mu_k)_k$ satisfy

$$\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_k] \geq (1-\epsilon)\mathrm{Var}_{\mu_k}[\phi]$$

for given $\epsilon > 0, k = 0, \dots, m-1$ and any integrable function $\phi : \mathscr{X} \to \mathbb{R}$, we have the telescoping product

$$\frac{\mathbb{E}[\mathrm{Var}_{\mu_m}[\phi]]}{\mathrm{Var}_\mu[\phi]} = \mathbb{E}\left[\prod_{k=0}^{m-1}\frac{\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_k]}{\mathrm{Var}_{\mu_k}[\phi]}\right] \geq (1-\epsilon)^m.$$

Hence, we have the bound

$$\mathfrak{gap}(K)^{-1} \leq (1-\epsilon)^{-m}$$

which immediately provides an upper bound for the mixing time of K in light of Theorem 2. This motivates the following definition.

Definition 3.6 (Discrete time approximate conservation of variance (ACV), [CE22]). A prelocalization process $(\mu_k)_k$ is said to conserve (η_k)-variance up to time m if for any integrable function $\phi : \mathscr{X} \to \mathbb{R}$, $0 \le k < m$,

$$\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_t] \ge (1-\eta_k)\mathrm{Var}_{\mu_k}[\phi].$$

By the same computation above, if $(\mu_k)_k$ conserves (η_k)-variance up to time m, then it associated dynamics at time m has a spectral gap of at least $\prod_{k=0}^{m-1}(1-\eta_k)$.

3.2.1 ACV of discrete time linear-tilt localization

In the case of the discrete time linear-tilt localization $(\mu_k)_k$ of some μ, we may compute an explicit value (η_k) for which $(\mu_k)_k$ conserves (η_k)-variance. Indeed, given integrable $\phi : \mathscr{X} \to \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_k] &= \mathbb{E}\left[\int \phi^2 \mathrm{d}\mu_{k+1} - \left(\int \phi \,\mathrm{d}\mu_{k+1}\right)^2 \,\middle|\, \mu_k\right] \\
&= \int \phi^2 \mathrm{d}\mu_k - \mathbb{E}\left[\left(\int \phi(x)(1+\langle x-\bar{\mu}_k, Z_k\rangle)\mu_k(\mathrm{d}x)\right)^2 \,\middle|\, \mu_k\right] \\
&= \int \phi^2 \mathrm{d}\mu_k - \left(\int \phi \,\mathrm{d}\mu_k\right)^2 - \mathbb{E}\left[\left(\int \phi(x)\langle x-\bar{\mu}_k, Z_k\rangle\mu_k(\mathrm{d}x)\right)^2 \,\middle|\, \mu_k\right] \\
&= \mathrm{Var}_{\mu_k}[\phi] - \mathrm{Var}\left[\int \phi(x)\langle x-\bar{\mu}_k, Z_k\rangle\mu_k(\mathrm{d}x) \,\middle|\, \mu_k\right]
\end{aligned} \tag{14}$$

where the third and last equality follows as $\mathbb{E}[Z_k \mid \mu_k] = 0$. Then, denoting

$$C_k := \mathrm{Cov}[Z_k \mid \mu_k] \text{ and } U_k := \int \phi(x)(x-\bar{\mu}_k)\mu_k(\mathrm{d}x),$$

Equation (14) becomes

$$\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_k] = \mathrm{Var}_{\mu_k}[\phi] - \|C_k^{1/2}U_k\|^2.$$

With this reduction, our goal is reduced to bounding $\|C_k^{1/2}U_k\|^2$ by an expression of the form $\eta_k \mathrm{Var}_{\mu_k}[\phi]$ where η_k is independent of ϕ. Indeed, taking θ to be the normalized vector $C^{1/2}U_k/\|C_k^{1/2}U_k\|$, we have

$$\begin{aligned}
\|C_k^{1/2}U_k\|^2 = \langle C_k^{1/2}U_k, \theta\rangle^2 &= \left(\int \phi(x)\langle C_k^{1/2}(x-\bar{\mu}_k), \theta\rangle\mu_k(\mathrm{d}x)\right)^2 \\
&= \left(\int \left(\phi(x) - \int \phi \,\mathrm{d}\mu_k\right)\langle C_k^{1/2}(x-\bar{\mu}_k), \theta\rangle\mu_k(\mathrm{d}x)\right)^2 \\
&\le \left(\int \left(\phi(x) - \int \phi \,\mathrm{d}\mu_k\right)^2 \mu_k(\mathrm{d}x)\right)\int \langle C_k^{1/2}(x-\bar{\mu}_k), \theta\rangle^2\mu_k(\mathrm{d}x) \\
&= \mathrm{Var}_{\mu_k}[\phi]\int \langle \theta, C_k^{1/2}(x-\bar{\mu}_k)^{\otimes 2}C_k^{1/2}\theta\rangle\mu_k(\mathrm{d}x) = \mathrm{Var}_{\mu_k}[\phi]\langle \theta, C_k^{1/2}\mathrm{Cov}(\mu_k)C_k^{1/2}\theta\rangle \\
&\le \mathrm{Var}_{\mu_k}[\phi]\|C_k^{1/2}\mathrm{Cov}(\mu_k)C_k^{1/2}\|_{\mathrm{op}}
\end{aligned}$$

where the inequality is due to Cauchy-Schwarz. Hence, for any discrete linear-tilt localization $(\mu_k)_k$, we have

$$\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_k] \geq (1 - \|C_k^{1/2}\mathrm{Cov}(\mu_k)C_k^{1/2}\|_{\mathrm{op}})\mathrm{Var}_{\mu_k}[\phi] \tag{15}$$

for any integrable $\phi : \mathscr{X} \to \mathbb{R}$ and thus, conserve $(1 - \eta_k)$-variance where we define $\eta_k := \|C_k^{1/2}\mathrm{Cov}(\mu_k)C_k^{1/2}\|_{\mathrm{op}}$.

3.2.2 ACV of linear-tilt localization driven by a Wiener process

In the setting of the linear-tilt localization driven by a Wiener process, a similar computation leveraging on the Cauchy-Schwarz inequality yields a useful inequality which turns out to be very useful with regards to the KLS conjecture. We will present the computation now while leaving its application to the KLS conjecture to Section 4.

In the continuous time setting, it no longer made sense to analyze the difference of the variance between subsequent time steps. Instead, we will analyze its time derivative. In particular, recalling the definition of the martingale (M_t) from Corollary 2.7, we may use the same method as the discrete time analysis to compute a bound for $\mathrm{d}[M_t]$.

Proposition 3.3. Given the linear-tilt localization $(\mu_t)_{t\geq 0}$ of μ driven by a Wiener process as defined in Section 2.1.1, for all isotropic $\phi : \mathbb{R}^n \to \mathbb{R}$, we have

$$\mathrm{d}[M]_t \leq \mathrm{Var}_{\mu_t}[\phi]\|A_t\|_{\mathrm{op}}\mathrm{d}t,$$

where $A_t := \int (x - a_t)^{\otimes 2}\mathrm{d}\mu_t$ is the covariance matrix of μ_t and $M_t := \int \phi\,\mathrm{d}\mu_t$.

Before proving this proposition, we first recall that for a martingale $(X_t)_t$, its quadratic variation $[X]_t$ is the unique adapted, continuous and non-decreasing process such that $(X_t^2 - [X]_t)_t$ is a martingale. In particular, if $X_0 = 0$, we observe

$$\mathbb{E}[X_t^2] - \mathbb{E}[X]_t = \mathbb{E}[X_0^2] = 0$$

for all t. Thus,

$$\mathrm{Var}[X_t] = \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2 = \mathbb{E}[X]_t - \mathbb{E}[X_0]^2 = \mathbb{E}[X]_t.$$

Now, since in our case ϕ is isotropic, we have $M_0 = \int \phi\,\mathrm{d}\mu = 0$ and so, $\mathrm{Var}[M_t] = \mathbb{E}[M]_t$. Thus, by taking expectation on both sides, one observes the resemblance of the above equation with that of the discrete time ACV inequality.

Proof. We compute

$$\begin{aligned}\mathrm{d}M_t &= d\int \phi(x)F_t(x)\mu(\mathrm{d}x) = \int \phi(x)\langle x - a_t, \mathrm{d}W_t\rangle\mu_t(\mathrm{d}x) \\ &= \left\langle \int \phi(x)(x - a_t)\mu_t(\mathrm{d}x), \mathrm{d}W_t \right\rangle = \langle U_t, \mathrm{d}W_t\rangle\end{aligned}$$

where we denote $U_t := \int \phi(x)(x - a_t)\mu_t(\mathrm{d}x)$ as in the discrete time case. Hence, by considering the component-wise quadratic variation, we have

$$\mathrm{d}[M]_t = \|U_t\|^2\,\mathrm{d}t. \tag{16}$$

The remainder of the calculation is quite similar to the discrete case. Indeed, denoting θ the vector U_t normalized to have norm 1, we obtain,

$$
\begin{aligned}
\mathrm{d}[M]_t = \langle U_t, \theta\rangle^2 \,\mathrm{d}t &= \left\langle \int \left(\phi(x) - \int \phi \,\mathrm{d}\mu_t\right)(x - a_t)\mu_t(\mathrm{d}x), \theta \right\rangle^2 \mathrm{d}t \\
&= \left(\int \left(\phi(x) - \int \phi \,\mathrm{d}\mu_t\right)\langle x - a_t, \theta\rangle \mu_t(\mathrm{d}x)\right)^2 \mathrm{d}t \\
&\leq \mathrm{Var}_{\mu_t}[\phi] \left(\int \langle \theta, (x - a_t)^{\otimes 2}\theta\rangle \mu_t(\mathrm{d}x)\right) \mathrm{d}t \\
&= \mathrm{Var}_{\mu_t}[\phi]\langle \theta, A_t\theta\rangle \mathrm{d}t \leq \mathrm{Var}_{\mu_t}[\phi] \|A_t\|_{\mathrm{op}} \mathrm{d}t.
\end{aligned}
\tag{17}
$$

where the second equality by applying Fubini's theorem and noting that $\int (x - a_t)\mu_t(\mathrm{d}x) = 0$. Moreover, the first inequality follows again by the Cauchy-Schwarz inequality. □

3.3 Glauber dynamics as an associated Markov process

Glauber dynamics can be constructed as an associated Markov chain of the coordinate by coordinate localization. Namely, taking $\mu \in \mathscr{M}(\{-1,1\}^n)$ and $\tau = n - 1$, we will show the associate Markov kernel

$$
\mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)}\right] = K(\sigma^1, \sigma^2)
$$

for all $\sigma^1, \sigma^2 \in \{-1,1\}^n$ where $K(\sigma^1, \sigma^2)$ as defined by Equation (10).

First, taking $\sigma^1, \sigma^2 \in \{-1,1\}^n$ such that $\|\sigma^1 - \sigma^2\| = 1$, say σ^1 and σ^2 differs at the m-th coordinate, we have

$$
\mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)}\right] = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)} \;\middle|\; Y_n = k\right].
$$

Now, if $Y_n = k \neq m$, the configuration is fixed at the m-th coordinate. However, as σ^1 and σ^2 differs at the m-th coordinate, either $\mu_{n-1}(\sigma^1) = 0$ or $\mu_{n-1}(\sigma^2) = 0$. Thus, all but the m-th term in the above sum vanishes and we have the kernel equals

$$
\begin{aligned}
&= \frac{1}{n}\mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)} \;\middle|\; Y_n = m\right] \\
&= \frac{1}{n}\mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)} \mathbf{1}_{\{\mathrm{supp}\ \mu_{n-1} = \{\sigma^1, \sigma^2\}\}} \;\middle|\; Y_n = m\right] \\
&= \frac{\mathbb{P}(\mathrm{supp}\ \mu_{n-1} = \{\sigma^1, \sigma^2\})}{n}\mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)} \;\middle|\; Y_n = m, \mathrm{supp}\ \mu_{n-1} = \{\sigma^1, \sigma^2\}\right]
\end{aligned}
$$

Then, substituting in the equalities,

$$
\mathbb{P}(\mathrm{supp}\ \mu_{n-1} = \{\sigma^1, \sigma^2\}) = \mu(\sigma^1) + \mu(\sigma^2),
$$

and

$$\begin{aligned}
&\mathbb{E}[\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2) \mid Y_n = m, \operatorname{supp} \mu_{n-1} = \{\sigma^1, \sigma^2\}] \\
=&\mathbb{E}[\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2) \mid \operatorname{supp} \mu_{n-1} = \{\sigma^1, \sigma^2\}] \\
=&\frac{\mu(\sigma^1)}{\mu(\sigma^1)+\mu(\sigma^2)} \frac{\mu(\sigma^2)}{\mu(\sigma^1)+\mu(\sigma^2)}
\end{aligned}$$

we have

$$\mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)}\right] = \frac{\mu(\sigma^1)+\mu(\sigma^2)}{n\mu(\sigma^1)} \frac{\mu(\sigma^1)}{\mu(\sigma^1)+\mu(\sigma^2)} \frac{\mu(\sigma^2)}{\mu(\sigma^1)+\mu(\sigma^2)} = \frac{1}{n} \frac{\mu(\sigma^2)}{\mu(\sigma^1)+\mu(\sigma^2)}$$

which is precisely the kernel of the Glauber dynamics of two neighboring configurations. On the other hand, if $\|\sigma^1 - \sigma^2\| > 1$, then

$$\mathbb{E}\left[\frac{\mu_{n-1}(\sigma^1)\mu_{n-1}(\sigma^2)}{\mu(\sigma^1)}\right] = 0$$

as μ_{n-1} have fixed all but one coordinate. Thus, the dynamics associated with the coordinate is precisely the Glauber dynamics (we don't need to check the case $\sigma^1 = \sigma^2$ as they are kernels and so $K(\sigma,\sigma) = 1 - \sum_{\tilde{\sigma}\neq\sigma} K(\sigma,\tilde{\sigma})$).

Now, as the coordinate by coordinate localization is a discrete linear-tilt localization as presented by Section 2.1.2, we have by Equation (15) that the spectral gap of the Glauber dynamics is bounded below by

$$\mathfrak{gap}(K) \geq \prod_{k=0}^{n-2} (1 - \|C_k^{1/2} \operatorname{Cov}(\mu_k) C_k^{1/2}\|_{\mathrm{op}}), \tag{18}$$

where $C_k := \operatorname{Cov}[Z_k \mid \mu_k]$ with Z_k defined by Equation (9). Furthermore, by observing from definition of Z_k that $(Z_k)_i(Z_k)_j = 0$ for all $i \neq j$, C_k has only non-zero entries on its diagonal for which

$$\begin{aligned}
(C_k)_{i,i} &= (\operatorname{Cov}[Z_k \mid \mu_k, Y_k = i])_{i,i} \mathbb{P}(Y_k = i) \\
&= \frac{1}{n-k}\left(\frac{1}{(1+(\bar{\mu}_k)_i)^2} \frac{1+(\bar{\mu}_k)_i}{2} + \frac{1}{(1-(\bar{\mu}_k)_i)^2} \frac{1-(\bar{\mu}_k)_i}{2}\right) \\
&= \frac{1}{n-k} \frac{1}{1-(\bar{\mu}_k)_i^2} = \frac{1}{n-k} (\operatorname{Cov}(\mu_k)_{i,i})^{-1}.
\end{aligned}$$

Thus, by introducing the correlation matrix $\operatorname{Cor}(\mu)$ with entries

$$\operatorname{Cor}(\mu)_{i,j} = \frac{\operatorname{Cov}(\mu)_{i,j}}{\sqrt{\operatorname{Cov}(\mu)_{i,i}}\sqrt{\operatorname{Cov}(\mu)_{j,j}}},$$

we have $C_k^{1/2} \operatorname{Cov}(\mu_k) C_k^{1/2} = \frac{1}{n-k} \operatorname{Cor}(\mu_k)$ and Equation (18) becomes

$$\mathfrak{gap}(K) \geq \prod_{k=0}^{n-2} \left(1 - \frac{\|\operatorname{Cor}(\mu_k)\|_{\mathrm{op}}}{n-k}\right). \tag{19}$$

This inequality allows us to directly recover a bound for the spectral gap should μ be spectrally independent. Indeed, denoting the maximum eigenvalue of a matrix A by $\rho(A)$, as $\mathrm{Cor}(\mu_k)$ is symmetric, we have

$$\|\mathrm{Cor}(\mu_k)\|_{\mathrm{op}} = \rho(\mathrm{Cor}(\mu_k)).$$

Furthermore, by observing

$$\Psi(\mu_k) = \mathrm{Cov}(\mu_k)\mathrm{diag}(\mathrm{Cov}(\mu_k))^{-1} - \mathrm{Id}_n,$$

we have

$$\|\mathrm{Cor}(\mu_k)\|_{\mathrm{op}} = \rho(\mathrm{Cov}(\mu_k)\mathrm{diag}(\mathrm{Cov}(\mu_k))^{-1}) = \rho(\Psi(\mu_k) + \mathrm{Id}_n) = \rho(\Psi(\mu_k)) + 1.$$

Thus, if μ is a measure on $\{-1,1\}^n$ which is $(\eta_0, \dots, \eta_{n-2})$-spectrally independent, we have $\rho(\Psi(\mu_k)) \leq \eta_k$ for all $k = 0, \dots, n-2$ and so, by Equation (18), we have

$$\mathfrak{gap}(K) \geq \prod_{k=0}^{n-2} \left(1 - \frac{\eta_k + 1}{n-k}\right). \tag{20}$$

This bound and Theorem 6 are said to be equivalent as both equations provide polynomial lower bounds for the spectral gap in terms of the spectral independence coefficients.

4 The KLS and Thin-Shell Conjecture

The stochastic localization scheme has been wildly successful in making progress towards the KLS conjecture. While Eldan introduced stochastic localization to prove Theorem 3, modifying the scheme slightly, Lee and Vempala in [LV16] had obtained the bound $C^n_{\text{con}} \lesssim n^{-1/4}$. Further modifying their arguments, a recent breakthrough by Chen in [Che20] improves the bound significantly by showing $C^n_{\text{con}} = n^{-o(1)}$. In this section, we present the original context for the construction of stochastic localizations. To this end, we will describe a proof of Eldan's original reduction of the KLS conjecture to the thin-Shell conjecture.

The method presented in this section is due to Lee and Vempala [LV16] and reformulated in the language of concentration by Eldan in [Eld18].

4.1 Some intuition for concentration

Heuristically, the concentration is a quantity which inversely measures how "spread out" the density of μ is in each direction. In the 1-dimensional case, the concentration of μ is therefore simply the inverse of its variance. This is illustrated by the following proposition.

Proposition 4.1. Let X be a $\mathbb{R}^n$-valued random variable. Then for all K-Lipschitz function $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\text{Var}[\phi(X)] \le K^2 \text{Var}[X].$$

Proof. We first prove the proposition in the case that $\mathbb{E}[X] = 0$.

Let X' be a i.i.d. copy of X on the same probability space. Then for all K-Lipschitz function ϕ, we have

$$\begin{aligned}
2\text{Var}[\phi(X)] &= \text{Var}[\phi(X) - \phi(X')] && \text{(i.i.d.)} \\
&= \mathbb{E}[(\phi(X) - \phi(X'))^2] - \mathbb{E}[\phi(X) - \phi(X')]^2 \\
&= \mathbb{E}[(\phi(X) - \phi(X'))^2] && \text{(identically distributed)} \\
&\le K^2 \mathbb{E}[\|X - X'\|^2] && \text{(as } \phi \text{ is } K\text{-Lipschitz)} \\
&= K^2 \mathbb{E}[X^T X + X'^T X' - X^T X' - X'^T X] \\
&= 2K^2 \text{Var}[X] - 2K^2 \text{Cov}(X, X') = 2K^2 \text{Var}[X]. && \text{(independence)}
\end{aligned}$$

implying $\text{Var}[\phi(X)] \le K^2 \text{Var}[X]$ as claimed.

For general X, by defining $\phi'(x) := \phi(x + \mathbb{E}[X])$, we can apply the 0 mean case to $X - \mathbb{E}X$ and ϕ' to obtain

$$\text{Var}[\phi(X)] = \text{Var}[\phi'(X - \mathbb{E}[X])] \le K^2 \text{Var}[X - \mathbb{E}[X]] = K^2 \text{Var}[X]$$

as required. □

With this proposition in mind, it is clear that for $\mathbb{R}$-valued random variables X, its law μ has concentration $C^\mu_{\text{con}} = \text{Var}[X]^{-1}$.

We note that the definition we are presenting here is slightly non-standard. However, utilizing the following remarkable result due to Milman, we show that this definition is equivalent to the following more standard definitions whenever we are working with log-concave measures.

Definition 4.1 (Exponential concentration, [Mil18])**.** Given a measure μ on $\mathbb{R}^n$, we say μ has exponential concentration if there exists some $c, D > 0$ such that for all 1-Lipschitz function $\phi : \mathbb{R}^n \to \mathbb{R}, t > 0$, we have

$$\mu(|\phi - \mathbb{E}_\mu[\phi]| \geq t) \leq ce^{-Dt}. \tag{21}$$

Fixing $c = 1$, we denote the largest possible D as D^μ_{exp}.

Definition 4.2 (First-moment concentration, [Mil18])**.** Again, taking μ a measure on $\mathbb{R}^n$, we say μ has first-moment concentration if there exists some $D > 0$ such that for all 1-Lipschitz function $\phi : \mathbb{R}^n \to \mathbb{R}$, we have

$$\mathbb{E}_\mu[|\phi - \mathbb{E}_\mu[\phi]|] \leq \frac{1}{D}. \tag{22}$$

We denote the largest possible D by D^μ_{FM}.

It is clear that exponential concentration implies first-moment concentration. Indeed, if μ has exponential concentration with constant D (taking $c = 1$), then by the tail probability formula,

$$\mathbb{E}_\mu[|\phi - \mathbb{E}_\mu[\phi]|] = \int_0^\infty \mu(|\phi - \mathbb{E}_\mu[\phi]| \geq t)\mathrm{d}t \leq \int_0^\infty e^{-Dt}\mathrm{d}t = \frac{1}{D}.$$

On the other hand, Milman showed that for log-concave measures on $\mathbb{R}^n$, exponential concentration and first-moment concentration are equivalent in the following sense.

Theorem 7 (Milman, [Mil08])**.** For all log-concave measure μ on $\mathbb{R}^n$, μ has exponential concentration if and only if μ has first-moment concentration. Furthermore, $D^\mu_{\text{exp}} \simeq D^\mu_{\text{FM}}$ where we write $A \simeq B$ if there exists universal constants $C_1, C_2 > 0$ such that $C_1 A \leq B \leq C_2 A$.

With this theorem in mind, we establish the following correspondence.

Proposition 4.2. For all measures μ on $\mathbb{R}^n$, we have

$$\text{Exponentially concentrated} \implies \text{Concentrated} \implies \text{First-moment concentrated}$$

and $D^\mu_{\text{exp}} \leq \sqrt{2}C^\mu_{\text{con}}$ and $C^\mu_{\text{con}}/2 \leq D^\mu_{\text{FM}}$. Hence, if μ is log-concave, $D^\mu_{\text{exp}} \simeq D^\mu_{\text{FM}} \simeq C^\mu_{\text{con}}$.

Proof. Assume first that μ is C-concentrated. Then by the Chebyshev inequality, we have

$$\mu(|\phi - \mathbb{E}_\mu[\phi]| \geq t) \leq \frac{1}{t^2}\mathrm{Var}_\mu[\phi] \leq \frac{1}{C^2 t^2},$$

for all 1-Lipschitz ϕ. Thus, by tail probability,

$$\begin{aligned}
\mathbb{E}_\mu[|\phi - \mathbb{E}_\mu[\phi]|] &= \int_0^\infty \mu(|\phi - \mathbb{E}_\mu[\phi]| \geq t)\mathrm{d}t \\
&\leq \inf_{a>0}\left\{\int_0^a \mu(|\phi - \mathbb{E}_\mu[\phi]| \geq t)\mathrm{d}t + \frac{1}{C^2}\int_a^\infty \frac{1}{t^2}\mathrm{d}t\right\} \\
&\leq \inf_{a>0}\left\{a + \frac{1}{aC^2}\right\} = \frac{2}{C},
\end{aligned}$$

implying μ is first-moment concentrated with respect to the constant $C/2$.

On the other hand, if μ is exponential concentrated with some constant D, then again by the tail probability,

$$\mathrm{Var}_\mu[\phi] = \int_0^\infty \mu((\phi - \mathbb{E}_\mu[\phi])^2 \geq t)\mathrm{d}t \leq \int_0^\infty e^{-D\sqrt{t}}\mathrm{d}t = \frac{2}{D^2}$$

implying μ is $D/\sqrt{2}$-concentrated. □

Thusly, for log-concave measures, all three notions of concentration presented above are equivalent.

4.2 Example: concentration of the Gaussian

As the main motivator for the KLS conjecture, the standard Gaussian measures on $\mathbb{R}^n$ satisfy a special property where its concentration is dimension invariant. This fact motivates the KLS conjecture which hypothesized that this invariance holds for all isotropic log-concave measures. We will for completeness give a brief proof (based on Theorem 1.7.1 in [Bog98]) of the following theorem taking $C = 2/\pi$ (although one can in addition show that the standard Gaussian measures are 1-concentrated).

Theorem 8 (Concentration of Gaussian measures)**.** Denoting γ^n the standard Gaussian measure on $\mathbb{R}^n$, γ^n is C-concentrated for some constant C which is independent of n. That is, for all 1-Lipschitz $\phi : \mathbb{R}^n \to \mathbb{R}$, $\mathrm{Var}_{\gamma^n}[\phi] \leq C^{-2}$.

To prove this theorem we first observe the following elementary property of the Gaussian measure.

Lemma 4.3. For all $x \in \mathbb{R}^n$, we have $\mathbb{E}_{\gamma^n}[|\langle x, \cdot\rangle|^2] = \|x\|^2$.

Proof. Defining $f_x := \langle x, \cdot \rangle$, we have $\mathbb{E}_{\gamma^n}[|\langle x, \cdot\rangle|^2] = \mathbb{E}_{(f_x)_*\gamma^n}[|\cdot|^2]$ where $(f_x)_*\gamma^n$ is the push-forward of γ^n along f_x. Since f_x is linear, it follows that $(f_x)_*\gamma^n \sim \mathcal{N}(0, \|x\|^2)$. Hence, the result follows as $\mathbb{E}_{(f_x)_*\gamma^n}[|\cdot|^2] = \mathrm{Var}_{(f_x)_*\gamma^n}[\mathrm{id}] = \|x\|^2$. □

With this in mind, fixing a smooth 1-Lipschitz function $\phi : \mathbb{R}^n \to \mathbb{R}$ (we can assume smoothness since any 1-Lipschitz function can be uniformly approximated by smooth 1-Lipschitz functions), we will now attempt to bound $\int\int |\phi(x) - \phi(y)|^2 \gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)$ by first bounding it by the integral of a inner product. In particular, for all $x, y \in \mathbb{R}^n$, we observe

$$\begin{aligned}
|\phi(x) - \phi(y)| &= \left| \int_0^{\pi/2} \partial_\theta \phi(x \sin\theta + y\cos\theta)\mathrm{d}\theta \right| \leq \int_0^{\pi/2} |\partial_\theta \phi(x\sin\theta + y\cos\theta)|\mathrm{d}\theta \\
&= \int_0^{\pi/2} |\langle \nabla\phi(x\sin\theta + y\cos\theta), x\cos\theta - y\sin\theta\rangle|\mathrm{d}\theta.
\end{aligned}$$

Then, by rescaling $\mathrm{d}\theta$, we may apply Jensen's inequality resulting in

$$\begin{aligned}
|\phi(x) - \phi(y)|^2 &\leq \left(\int_0^{\pi/2} |\langle \nabla\phi(x\sin\theta + y\cos\theta), x\cos\theta - y\sin\theta\rangle|\mathrm{d}\theta \right)^2 \\
&\leq \frac{\pi}{2} \int_0^{\pi/2} |\langle \nabla\phi(x\sin\theta + y\cos\theta), x\cos\theta - y\sin\theta\rangle|^2\mathrm{d}\theta
\end{aligned}$$

Thus, we have

$$\begin{aligned}&\int\int |\phi(x)-\phi(y)|^2\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)\\&\le \frac{\pi}{2}\int_0^{\pi/2}\mathrm{d}\theta\int\int |\langle\nabla\phi(x\sin\theta+y\cos\theta), x\cos\theta-y\sin\theta\rangle|^2\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y).\end{aligned}$$

Now, by substituting $u=x\sin\theta+y\cos\theta, v=x\cos\theta-y\sin\theta$ (which Jacobian has determinant 1), we have

$$\begin{aligned}\int\int |\phi(x)-\phi(y)|^2\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) &\le \frac{\pi}{2}\int_0^{\pi/2}\mathrm{d}\theta\int\int |\langle\nabla\phi(u), v\rangle|^2\gamma^n(\mathrm{d}u)\gamma^n(\mathrm{d}v)\\&\le \frac{\pi}{2}\int_0^{\pi/2}\mathrm{d}\theta\int \|\nabla\phi(u)\|^2\gamma^n(\mathrm{d}u)\end{aligned}$$

where the second inequality is due to Lemma 4.3. Hence, as ϕ is 1-Lipschitz, $\|\nabla\phi(u)\|\le 1$ for all u and thus,

$$\int\int |\phi(x)-\phi(y)|^2\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)\le \frac{\pi}{2}\int_0^{\pi/2}\mathrm{d}\theta\int\mathrm{d}\gamma^n=\frac{\pi}{2}\cdot\frac{\pi}{2}=\frac{\pi^2}{4}.$$

The immediate consequence of this bound is that there exists some $z\in\mathbb{R}^n$ such that $\int|\phi(x)-z|^2\gamma^n(\mathrm{d}x)\le\pi^2/4$. Indeed, if no such z exists, then $\int|\phi(x)-\phi(y)|^2\gamma^n(\mathrm{d}x)>\pi^2/4$ for all y, implying $\int\int|\phi(x)-\phi(y)|^2\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)>\pi^2/4$ which contradicts the above bound. Hence, choosing such a z, we conclude the proof of Theorem 8 since

$$\mathrm{Var}_{\gamma^n}[\phi]=\min_{w\in\mathbb{R}^n}\int|\phi(x)-w|^2\gamma^n(\mathrm{d}x)\le\int|\phi(x)-z|^2\gamma^n(\mathrm{d}x)\le\left(\frac{\pi}{2}\right)^2$$

as required.

4.3 Discussion of the KLS conjecture

4.3.1 Isotropic as a normalization condition

To address our concern that in general the concentration of log-concave measures are not invariant under linear maps, we observe that all isotropic measures μ satisfies $\mathrm{Var}_\mu[T]\le 1$ for all 1-Lipschitz linear maps $T:\mathbb{R}^n\to\mathbb{R}$. Indeed, if $T:\mathbb{R}^n\to\mathbb{R}$ is a 1-Lipschitz linear function, i.e. it is of the form $v\mapsto w^Tv+d$ for some $w\in S^{n-1}$ and $d\in\mathbb{R}$, then we have

$$\mathrm{Var}_\mu[T]=\mathrm{Var}_{X\sim\mu}\left[\sum_{i=1}^n w_iX_i+d\right]=\sum_{i,j=1}^n w_iw_j\mathrm{Cov}_{X\sim\mu}(X_i,X_j)=\sum_{i=1}^n w_i^2=1,$$

as $\mathrm{Cov}_{X\sim\mu}(X)=\mathrm{id}$.

With the above property, we remark a trivial bound on the concentration of μ: Taking $\pi_i:\mathbb{R}^n\to\mathbb{R}$ to be the projection into the i-th coordinate, as π_i is linear and 1-Lipschitz for all i, we have

$\mathrm{Var}_{X\sim\mu}[X_i] \le 1$. Thus,

$$\begin{aligned}\mathrm{Var}_{X\sim\mu}[X] &= \mathbb{E}_{X\sim\mu}[X^T X] - \mathbb{E}_{X\sim\mu}[X]^T \mathbb{E}_{X\sim\mu}[X] \\ &= \sum_{i=1}^{n} (\mathbb{E}_{X\sim\mu}[X_i^2] - (\mathbb{E}_{X\sim\mu} X_i)^2) = \sum_{i=1}^{n} \mathrm{Var}_\mu[\pi_i] \le n.\end{aligned}$$

Thus, by Proposition 4.1, we have that $\mathrm{Var}_\mu[\phi] \le n$ for all 1-Lipschitz $\phi : \mathbb{R}^n \to \mathbb{R}$ implying μ is $n^{-1/2}$-concentrated[1].

4.3.2 Equivalent formulation of the KLS conjecture

While we have formulated the KLS conjecture using the language of concentration, the conjecture can be instead formulated in several equivalent ways. As demonstrated by Proposition 4.2, it is clear that the KLS conjecture can be equivalently formulated using the exponential or first moment concentration. Moreover, two other common formulation of the KLS conjecture involves the Cheeger constant and the Poincaré inequality. For completeness of this exposition, we will now also briefly present the latter two equivalent formulations here.

The Cheeger constant is a quantity relating to the proportion of the volume of a set with its surface area. This quantity arises naturally when studying the isoperimetric problem. The isoperimetric problem is the problem in finding the set of unit volume with minimum surface area. In the case of $\mathbb{R}^n$ equipped with the Lebesgue measure, we have known since the ancient Greeks that the solution is the unit ball. With this in mind, the KLS conjecture can be viewed geometrically as a natural extension of the isoperimetric problem replacing the Lebesgue measure with general log-concave measures on $\mathbb{R}^n$.

Definition 4.3 (Minkowski's boundary measure)**.** Given a measure μ on $\mathbb{R}^n$ and a Borel set $A \subseteq \mathbb{R}^n$, the Minkowski's boundary measure of A is defined as

$$\mu^+(\partial A) := \limsup_{\epsilon \downarrow 0} \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon}.$$

where $A_\epsilon := \{x \in \mathbb{R}^n \mid \mathrm{dist}(x, A) \le \epsilon\}$ is the ϵ-thickening of A.

A similar definition for the boundary measure μ^- can be given by taking the liminf instead. This does not matter in our case as any convex subsets of $\mathbb{R}^n$ are regular and the two definitions coincide.

The isoperimetric problem for the measure μ then becomes the problem of finding the set A of a fixed volume with minimum boundary measure.

Definition 4.4 (Cheeger's inequality, [Mil08])**.** Given a probability measure μ on $\mathbb{R}^n$, we say μ satisfy Cheeger's inequality if there exists some D such that for all A with $\mu(A) \le 1/2$,

$$D\mu(A) \le \mu^+(A).$$

We call the largest such D the Cheeger's constant and denote it by D_C^μ.

With these definitions, the KLS conjecture can be equivalently reformulated as the following.

[1] This is also a consequence of Theorem 2 in [LV16].

Conjecture 4 (KLS, [Eld13]). Denoting as before $\mathscr{M}^n_{\text{iso}}$ the set of all log-concave and isotropic probability measures μ on $\mathbb{R}^n$, there exists a *universal* constant D such that for all $\mu \in \mathscr{M}^n_{\text{iso}}$, μ satisfy the Cheeger's inequality with constant D.

The equivalence of the reformulation follows by completing Theorem 4.2 with two additional equivalences.

Definition 4.5 (Poincaré inequality, [Mil08]). Given a probability measure μ on $\mathbb{R}^n$, we say μ satisfy the Poincaré inequality if there exists some $D > 0$ such that for all smooth $\phi : \mathbb{R}^n \to \mathbb{R}$ satisfying $\int \phi \mathrm{d}\mu = 0$, we have

$$D\mathrm{Var}_\mu[\phi] \leq \mathbb{E}_\mu[\|\nabla\phi\|^2] = \int \|\nabla\phi\|^2 \mathrm{d}\mu.$$

We denote the largest such D by D^μ_{P}.

Theorem 9 (Milman, [Mil08]). For all log-concave measure μ on $\mathbb{R}^n$, the following are equivalent.

- There exists some $D < \infty$ such that μ has exponential concentration with constant D. We denote the largest such D by D^μ_{exp}.
- There exists some $D < \infty$ such that μ has first-moment concentration with constant D. We denote the largest such D by D^μ_{FM}.
- There exists some $D < \infty$ such that μ satisfy the Cheeger's inequality with constant D. We denote the largest such D by D^μ_{C}.
- There exists some $D < \infty$ such that μ satisfy the Poincaré inequality with constant D. We denote the largest such D by D^μ_{P}.

Furthermore, $D^\mu_{\text{exp}} \simeq D^\mu_{\text{FM}} \simeq D^\mu_{\text{P}} \simeq (D^\mu_{\text{C}})^2$.

With Theorem 9 and Proposition 4.2 in mind, it is clear that the KLS conjecture can be instead formulated with any of these inequalities instead. It is also worthwhile to pay attention to the formulation using the Poincaré inequality as it is relevant to the subsequent section on the log-Sobolev inequality.

Conjecture 5 (KLS, [Eld13]). Denoting $\mathscr{M}^n_{\text{iso}}$ as above, there exist a *universal* constant D such that for all $\mu \in \mathscr{M}^n_{\text{iso}}$, μ satisfy the Poincaré inequality with constant D.

4.4 Reduction of KLS to thin-shell

We will now present a proof of Theorem 3. As a high level overview, recall that the linear-tilt localization of a given measure is a measure-valued martingale for which the original measure is recovered in the limit. Then, as the concentration of the measure relates to the covariance of said measure, we will stop the martingale before the covariance grows too large. This allows us to analyze the martingale in a more tractable manner. However, as the the sequence is a martingale, some properties are invariant in time and hence allowing us to conclude that these properties also hold for the original measure.

We recall the goal of Theorem 3 is to control $\mathrm{Var}_\mu[\phi]$ by a logarithmic factor of $\mathrm{Var}_\mu[\|\cdot\|]$. We will assume μ is supported on $B_n(0) \subseteq \mathbb{R}^n$ with $B_n(0)$ the ball at the origin of radius n. The reason for this is due to a concentration bound for log-concave measures where one may show most of their densities lie within a compact support.

Lemma 4.4 ([Kla06], Equation 10)**.** There exists a universal constant C such that for all $f : \mathbb{R}^n \to \mathbb{R}$ the density of a log-concave isotropic measure, we have

$$f(x) \le f(0)e^{-|x|/C},$$

for all $x \in \mathbb{R}^n$ with $\|x\| \ge Cn$.

As a result, the region outside of said ball contributes to the exponential concentration by at most

$$\begin{aligned}\mu(|\phi - \mathbb{E}_\mu[\phi]| \ge t \cap CB_n(0)) &\le \int_{\{|\phi(x)|\ge t\}\cap CB_n(0)} e^{-|x|/c}\mathrm{d} \\ &\simeq \int_{t\vee n} r^{n-1}e^{-r}\mathrm{d}r = O(e^{-t}).\end{aligned}$$

Namely, the exponential concentration constant of μ outside of the ball is universally bounded. Hence, it follows by Theorem 9 that it is sufficient to bound the concentration of μ within $B_n(0)$. Consequently, we also have

$$\mathrm{supp}\ \mu_t = \mathrm{supp}\ F_t\mu \subseteq \mathrm{supp}\ \mu \subseteq B_n(0)$$

for all $t > 0$.

Let us fix $\phi : \mathbb{R}^n \to \mathbb{R}$ some 1-Lipschitz function and let (M_t) be the martingale as described in Corollary 2.7, we have $\mathrm{Var}_\mu[\phi] = \mathrm{Var}[\phi(a_\infty)]$ where $a_\infty \sim \mu$. Then, for all $t > 0$, by the law of total variance and the martingale property we have

$$\begin{aligned}\mathrm{Var}_\mu[\phi] = \mathrm{Var}[M_\infty] &= \mathrm{Var}[\mathbb{E}[M_\infty \mid \mu_t]] + \mathbb{E}[\mathrm{Var}[M_\infty \mid \mu_t]] \\ &= \mathrm{Var}[M_t] + \mathbb{E}[\mathrm{Var}[M_\infty \mid \mu_t]]\end{aligned} \tag{23}$$

where we introduce the notation $\mathrm{Var}[X \mid \mathscr{G}] := \mathbb{E}[(X - \mathbb{E}[X \mid \mathscr{G}])^2 \mid \mathscr{G}]$ for some random variable X and sub-σ-algebra $\mathscr{G}$. Furthermore, by replacing ϕ in Corollary 2.7 with ϕ^2 and denoting the resulting martingale $N_t := \int \phi^2 \mathrm{d}\mu_t$, we obtain $N_t \to \phi(a_\infty)^2$ and hence,

$$\begin{aligned}\mathrm{Var}[M_\infty \mid \mu_t] &= \mathbb{E}[\phi(a_\infty)^2 \mid \mu_t] - M_t^2 = N_t - M_t^2 \\ &= \int \phi^2 \mathrm{d}\mu_t - \left(\int \phi\,\mathrm{d}\mu_t\right)^2 = \mathrm{Var}_{\mu_t}[\phi]\end{aligned}$$

Combining this with Equation (23), we obtain

$$\mathrm{Var}_\mu[\phi] = \mathrm{Var}[M_t] + \mathbb{E}[\mathrm{Var}_{\mu_t}[\phi]]. \tag{24}$$

for any $t \ge 0$. Furthermore, by applying the optional stopping theorem, the same equality holds when we take t to be a stopping time.

At this point, by recalling Proposition 3.3, we recognize that the first term $\mathrm{Var}[M_t]$ is controlled by the operator norm of A_t with A_t being the covariance matrix of μ_t. Thus, to bound $\mathrm{Var}[M_t]$, the idea is to choose an appropriate stopping time τ to stop the process before $\|A_t\|_{\mathrm{op}}$ grows too large. On the other hand, for the given τ, by plugging in Equation (5) into Lemma 2.4, the second term $\mathrm{Var}_{\mu_\tau}[\phi]$ is then bounded by τ^{-1} for which the expectation can be bounded explicitly. We dedicate the remainder of this section to describe said procedure in detail.

4.4.1 Analysis of the covariance matrix

As demonstrated in Section 2.1.1, we know the limiting behavior of the covariance matrices, namely $A_t \to 0$ entry-wise as $t \to \infty$. This was important for us to establish the existence of the limit of (a_t) and (M_t). However, as shown above, we now require some quantitative bounds for the operator norm of A_t. For this purpose, we first compute some useful properties of A_t.

Observing

$$\int \mathrm{d}F_t(x)\mu(\mathrm{d}x) = \int \langle x - a_t, \mathrm{d}W_t\rangle \mu_t(\mathrm{d}x) = \left\langle \int x\mu_t(\mathrm{d}x) - a_t, \mathrm{d}W_t \right\rangle = 0,$$

we have

$$\begin{aligned} \mathrm{d}a_t &= \mathrm{d}\int xF_t(x)\mu(\mathrm{d}x) = \int x\mathrm{d}F_t(x)\mu(\mathrm{d}x) = \int (x-a_t)\mathrm{d}F_t(x)\mu(\mathrm{d}x) \\ &= \int (x-a_t)\langle x-a_t, \mathrm{d}W_t\rangle F_t(x)\mu(\mathrm{d}x) = \int (x-a_t)^{\otimes 2}\mathrm{d}W_t\mu_t(\mathrm{d}x) = A_t\mathrm{d}W_t \end{aligned} \tag{25}$$

where the second to last equality used the fact that $v\langle v, w\rangle = v^{\otimes 2}w$ for any appropriate v, w.

Similarly, computing using Itô's formula, we have

$$\begin{aligned} \mathrm{d}A_t &= \mathrm{d}\int (x-a_t)^{\otimes 2}F_t(x)\mu(\mathrm{d}x) \\ &= \int (x-a_t)^{\otimes 2}\mathrm{d}F_t(x) + F_t(x)\,\mathrm{d}(x-a_t)^{\otimes 2} \\ &\qquad - 2(x-a_t)\otimes \mathrm{d}[a_t, F_t(x)]_t + F_t(x)\mathrm{d}[a_t]_t\mu(\mathrm{d}x). \end{aligned} \tag{26}$$

The second term vanishes as

$$\int F_t(x)\,\mathrm{d}(x-a_t)^{\otimes 2}\,\mu(\mathrm{d}x) = -2\mathrm{d}a_t \otimes \overbrace{\int (x-a_t)\mu_t(\mathrm{d}x)}^{=0} = 0.$$

Also, by Equation (25), $\mathrm{d}a_t = A_t\mathrm{d}W_t$ implying $\mathrm{d}[a_t]_t = A_t^2\mathrm{d}t$. Finally, as both (a_t) and $(F_t(x))$

are martingales, $\mathrm{d}[a_t, F_t(x)]_t = F_t(x)A_t x\mathrm{d}t$ and the third term becomes

$$
\begin{aligned}
-2\int (x-a_t)\otimes \mathrm{d}[a_t, F_t(x)]\mu_t(\mathrm{d}x) &= -2A_t\left(\int (x-a_t)\otimes x\mu_t(\mathrm{d}x)\right)\mathrm{d}t \\
&= -2A_t\left(\overbrace{\int (x-a_t)^{\otimes 2}\mu_t(\mathrm{d}x)}^{A_t} + \overbrace{\int (x-a_t)\mu_t(\mathrm{d}x)}^{=0}\otimes a_t\right)\mathrm{d}t \\
&= -2A_t^2\mathrm{d}t.
\end{aligned}
$$

Hence, combining these and Equation (3) together in (26), we have

$$
\mathrm{d}A_t = \int (x-a_t)^{\otimes 2}\langle x-a_t, \mathrm{d}W_t\rangle\mu_t(\mathrm{d}x) - A_t^2\mathrm{d}t
$$

However, since we wish to bound A_t from above, as the drift term $-A_t^2\mathrm{d}t$ only contributes negatively, an upper bound for the process of the form $\int (x-a_t)^{\otimes 2}\langle x-a_t, \mathrm{d}W_t\rangle\mu_t(\mathrm{d}x)$ is also sufficient for A_t. Hence, we proceed by ignoring the drift term and redefine the process A_t such that

$$
\mathrm{d}A_t = \int (x-a_t)^{\otimes 2}\langle x-a_t, \mathrm{d}W_t\rangle\mu_t(\mathrm{d}x). \tag{27}
$$

With this justification, we now proceed to bound the operator norm of this new A_t. In particular, as A_t is symmetric, we recall that

$$
\|A_t\|_{\mathrm{op}} = \max_{i=1,\ldots,n}\lambda_i(t) = \|(\lambda_i(t))_{i=1}^n\|_\infty
$$

where $\lambda_i(t)$ denotes the distinct eigenvalues of A_t. Hence, it suffices to find a bound for the potential

$$
\Phi^\alpha(t) = \sum_{i=1}^n |\lambda_i(t)|^\alpha = \|(\lambda_i(t))_{i=1}^n\|_\alpha^\alpha \tag{28}
$$

for some $\alpha > 0$. Furthermore, as A_t is positive semi-definite, $\lambda_i(t) \geq 0$ for all $i = 1, \ldots, n$ and thus we have $\Phi^\alpha(t) = \sum_{i=1}^n \lambda_i(t)^\alpha$. Thus, by utilizing the following simple lemma, we can apply Itô's formula to establish a SDE of the form

$$
\mathrm{d}\Phi^\alpha(t) \lesssim 2\alpha^2 K_t \Phi^\alpha(t)\mathrm{d}t + \text{martingale}
$$

which will in turn allow us to bound Φ^α using Gronwall's inequality.

Lemma 4.5. If $A = [a_{ij}]$ is a diagonal matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then for all $i, j, k, l, m \in 1, \ldots n$, we have

- $\frac{\partial \lambda_i}{\partial a_{jk}} = \delta_{ij}\delta_{ik}$;
- whenever $i \neq j$, $\frac{\partial^2 \lambda_i}{\partial a_{ij}^2} = 2(\lambda_i - \lambda_j)^{-1}$;
- and for $j \neq l, k \neq m$ or $i \neq j$ and $i \neq k$, $\frac{\partial^2 \lambda_i}{\partial a_{jk}\partial a_{lm}} = 0$,

where δ_{ij} denotes the Kronecker delta function.

As this lemma requires the matrix to be diagonal, denoting $e_1, \dots, e_n$ as the normalized eigenbasis of A_{t_0} (they are in fact orthonormal as A_{t_0} is positive semi-definite), we will consider A_t with respect to this basis by considering the entries

$$a_{ij}(t) := \langle e_i, A_t e_j \rangle.$$

Using Equation (27), we compute

$$\begin{aligned} \mathrm{d}a_{ij}(t) &= \left\langle e_i, \left(\int (x - a_t)^{\otimes 2} \langle x - a_t, \mathrm{d}W_t \rangle \mu_t(\mathrm{d}x) \right) e_j \right\rangle \\ &= \left\langle \int \langle e_i, (x - a_t)^{\otimes 2} e_j \rangle (x - a_t) \mu_t(\mathrm{d}x), \mathrm{d}W_t \right\rangle = \langle \xi_{ij}, \mathrm{d}W_t \rangle \end{aligned}$$

where we introduce the notation

$$\xi_{ij}(t) = \int \langle e_i, (x - a_t)^{\otimes 2} e_j \rangle (x - a_t) \mu_t(\mathrm{d}x).$$

Thus, combining this with Lemma 4.5, denoting $\lambda_i = \lambda_i(t_0)$, we have by Itô's formula

$$\begin{aligned} \mathrm{d}\lambda_i(t) &= \sum_{j,k=1}^{n} \frac{\partial \lambda_i}{\partial a_{jk}} \mathrm{d}a_{jk}(t) + \frac{1}{2} \sum_{j,k=1}^{n} \sum_{l,m=1}^{n} \frac{\partial^2 \lambda_i}{\partial a_{jk} \partial a_{lm}} \mathrm{d}[a_{jk}, a_{lm}]_t \\ &= \langle \xi_{ii}, \mathrm{d}W_t \rangle + \sum_{j \neq i} \frac{\mathrm{d}[a_{ij}]_t}{\lambda_i - \lambda_j} = \langle \xi_{ii}, \mathrm{d}W_t \rangle + \sum_{j \neq i} \frac{\|\xi_{ij}\|^2}{\lambda_i - \lambda_j} \mathrm{d}t. \end{aligned} \tag{29}$$

at $t = t_0$. As a result, it is also clear that $\mathrm{d}[\lambda_i]_{t_0} = \|\xi_{ii}\|^2 \mathrm{d}t$.
Again applying Itô's formula, we have that

$$\begin{aligned} \mathrm{d}\Phi^\alpha(t) &= \sum_{i=1}^{n} \frac{\partial \Phi^\alpha}{\partial \lambda_i}\Big|_{t=t_0} \mathrm{d}\lambda_i(t) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 \Phi^\alpha}{\partial \lambda_i \partial \lambda_j}\Big|_{t=t_0} \mathrm{d}[\lambda_i, \lambda_j]_t \\ &= \alpha \sum_{i=1}^{n} \lambda_i^{\alpha-1} \mathrm{d}\lambda_i(t) + \frac{1}{2}\alpha(\alpha-1) \sum_{i=1}^{n} \lambda_i^{\alpha-2} \mathrm{d}[\lambda_i]_t \\ &= \alpha \sum_{i=1}^{n} \lambda_i^{\alpha-1} \left(\langle \xi_{ii}, \mathrm{d}W_t \rangle + \sum_{j \neq i} \frac{\|\xi_{ij}\|^2}{\lambda_i - \lambda_j} \mathrm{d}t \right) + \frac{1}{2}\alpha(\alpha-1) \sum_{i=1}^{n} \lambda_i^{\alpha-2} \mathrm{d}[\lambda_i]_t \\ &= \alpha \sum_{i \neq j} \lambda_i^{\alpha-1} \frac{\|\xi_{ij}\|^2}{\lambda_i - \lambda_j} \mathrm{d}t + \frac{1}{2}\alpha(\alpha-1) \sum_{i=1}^{n} \lambda_i^{\alpha-2} \|\xi_{ii}\|^2 \mathrm{d}t + \Bigg\langle \underbrace{\alpha \sum_{i=1}^{n} \lambda_i^{\alpha-1} \xi_{ii}}_{=:v_t}, \mathrm{d}W_t \Bigg\rangle \\ &= \frac{1}{2}\alpha \sum_{i \neq j} \|\xi_{ij}\|^2 \frac{\lambda_i^{\alpha-1} - \lambda_j^{\alpha-1}}{\lambda_i - \lambda_j} \mathrm{d}t + \frac{1}{2}\alpha(\alpha-1) \sum_{i=1}^{n} \lambda_i(t)^{\alpha-2} \|\xi_{ii}\|^2 \mathrm{d}t + \langle v_t, \mathrm{d}W_t \rangle. \end{aligned}$$

Now, since

$$\frac{\lambda_i^{\alpha-1} - \lambda_j^{\alpha-1}}{\lambda_i - \lambda_j} = \lambda_i^{\alpha-2} + \lambda_i^{\alpha-3}\lambda_j + \cdots + \lambda_i^{\alpha-2} \le (\alpha-1)(\lambda_i \vee \lambda_j)^{\alpha-2}.$$

the above expression is then bounded by

$$
\begin{aligned}
\mathrm{d}\Phi^\alpha(t) &\le \frac{1}{2}\alpha(\alpha-1)\sum_{i\neq j}\|\xi_{ij}\|^2(\lambda_i \vee \lambda_j)^{\alpha-2}\mathrm{d}t + \frac{1}{2}\alpha(\alpha-1)\sum_{i=1}^n \lambda_i(t)^{\alpha-2}\|\xi_{ii}\|^2\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle \\
&= \frac{1}{2}\alpha(\alpha-1)\sum_{i,j=1}^n \|\xi_{ij}\|^2(\lambda_i \vee \lambda_j)^{\alpha-2}\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle \\
&\le \alpha^2 \sum_{i,j=1}^n \|\xi_{ij}\|^2 \lambda_i^{\alpha-2}\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle,
\end{aligned}
$$

Thus, we have shown

$$
\mathrm{d}\Phi^\alpha(t) \le \alpha^2 \sum_{i=1}^n \lambda_i(t)^{\alpha-2}\sum_{j=1}^n \|\xi_{ij}\|^2\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle \tag{30}
$$

where $v_t := \alpha \sum_{i=1}^n \lambda_i^{\alpha-1}\xi_{ii}$.

By recalling that our goal is to bound $\|A_t\|_{\mathrm{op}}$ from above (c.f. Equation (24) and (17)), we may assume without loss of generality that $\|A_t\|_{\mathrm{op}} \ge 1$. Thus, applying the reverse Cauchy-Schwarz inequality to Equation (30), we have

$$
\begin{aligned}
\mathrm{d}\Phi^\alpha(t) &\le 2\alpha^2 \sum_{i=1}^n \lambda_i(t)^{\alpha-2}\sum_{j=1}^n \|\xi_{ij}\|^2\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle \\
&\le 2\alpha^2\|A_t\|_{\mathrm{op}}^2 \sum_{i=1}^n \lambda_i(t)^{\alpha-2}\sum_{j=1}^n \|\xi_{ij}\|^2\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle \\
&\lesssim 2\alpha^2 \sum_{i=1}^n \lambda_i(t)^{\alpha}\sum_{j=1}^n \|\xi_{ij}\|^2\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle.
\end{aligned}
$$

Thus, defining $K_t := \sup_i \sum_{j=1}^n \|\xi_{ij}(t)\|^2$, we have the desired asymptotic bound of the form

$$
\mathrm{d}\Phi^\alpha(t) \lesssim 2\alpha^2 K_t \Phi^\alpha(t)\mathrm{d}t + \langle v_t, \mathrm{d}W_t\rangle. \tag{31}
$$

4.4.2 Stopping the process early

As outlined in the beginning of this section, we will stop the process early in order to provide a bound for the right hand side of Equation (24). By observing Equation (17), we hypothesize that we should stop the process once $\|A_t\|_{\mathrm{op}}$ grows too large. As a result we define the stopping time

$$
\tau := \inf\{t > 0 \mid \|A_t\|_{\mathrm{op}} > 2\} \wedge 1.
$$

By the optional stopping theorem, Proposition 3.3 and the fact that we had assumed μ is supported in a ball of radius n, we have

$$
\begin{aligned}
[M]_\tau = \int_0^\tau \mathrm{d}[M]_t &\le \int_0^\tau \overbrace{\mathrm{Var}_{\mu_t}[\phi]}^{\le t^{-1}\wedge n^2}\overbrace{\|A_t\|_{\mathrm{op}}}^{\le 2}\mathrm{d}t \\
&\le 2\int_0^\tau t^{-1}\wedge n^2\mathrm{d}t \le 2\int_0^1 t^{-1}\wedge n^2\mathrm{d}t = 2 + 4\log n.
\end{aligned}
$$

Combining this with Equation (24) (where we recall that $\mathrm{Var}[M_t] = \mathbb{E}[M]_t$), we obtain

$$\mathrm{Var}_\mu[\phi] \leq 2 + 4\log n + \mathbb{E}[\tau^{-1}], \tag{32}$$

and it remains to find an upper bound for $\mathbb{E}[\tau^{-1}]$. Observing that $t < \tau$ whenever $\Phi^\alpha(t) < 2^\alpha$, we define σ the first time for which the potential $\Phi^\alpha(t)$ reaches 2^α, namely

$$\sigma := \inf\{t > 0 \mid \Phi^\alpha(t) = 2^\alpha\},$$

we have $\sigma^{-1} \geq \tau^{-1}$ and so it suffices to bound σ from below.

For simplicity, we will ignore the martingale term in Equation (31) and regard it as an ODE so that we can apply Gronwall's inequality.

Lemma 4.6 (Gronwall's inequality)**.** Let $T > 0$ and f a non-negative, bounded, measurable function on $[0, T]$ such that there exists some $b \geq 0$ for which,

$$\mathrm{d}f(t) \leq bf(t)\mathrm{d}t, \ \forall t \in [0, T].$$

Then, $f(t) \leq f(0)e^{bt}$ for all $t \in [0, T]$.

Hence, if we can find some constant K such that $K_t \leq K$ for all $t \leq \tau$, as $\Phi^\alpha(0) = n$, we have the bound

$$\Phi^\alpha(t) \leq ne^{2\alpha^2 Kt}.$$

Thus, substituting σ into the above, we have

$$2^\alpha = \Phi^\alpha(\sigma) \leq ne^{2\alpha^2 K\sigma}$$

implying

$$\frac{\alpha\log 2 - \log n}{2\alpha^2 K} \leq \sigma \leq \tau.$$

Then, taking $\alpha = 10K\log n$, it is easy to check that

$$\frac{1}{10K\log n} \leq \frac{\alpha\log 2 - \log n}{2\alpha^2 K}$$

implying $\mathbb{E}[\tau^{-1}] \leq 10K\log n$.

Whilst this deduction only holds when ignoring the stochastic term $\langle v_t, \mathrm{d}W_t\rangle$, one can show that the stochastic term grows much slower than the drift term and hence can be safely ignored. Indeed, it is possible to show that (c.f. second part of Lemma 34 in [LV18])

$$\langle v_t, \mathrm{d}W_t\rangle \lesssim \alpha\Phi^\alpha(t)\mathrm{d}W_t.$$

Namely, the martingale term grows at a rate of at most $\alpha\Phi^\alpha(t)\sqrt{t}$ which is much slower than that of the drift term.

Finally, to find a bound for (K_t), we employ the following lemma.

Lemma 4.7 (Lemma 1.6 in [Eld13])**.** Denoting C_{TS}^n as in Theorem 3, there exists a constant C such that for any log-concave, isotropic probability measure μ, we have

$$\sup_{\theta\in S^{n-1}} \sum_{i,j=1}^n \mathbb{E}_{X\sim\mu}[\langle X, e_i\rangle\langle X, e_j\rangle\langle X, \theta\rangle]^2 \leq C\sum_{k=1}^n \frac{1}{k(C_{\mathrm{TS}}^n)^2},$$

where $\{e_1, \ldots, e_n\}$ is any orthonormal basis on $\mathbb{R}^n$.

Recalling that

$$\xi_{ij} = \mathbb{E}_{X+a_t\sim\mu_t}[\langle e_i, X^{\otimes 2} e_j\rangle X] = \mathbb{E}_{X+a_t\sim\mu_t}[\langle X, e_i\rangle\langle X, e_j\rangle X],$$

we have by Parseval's identity

$$\begin{aligned}
K_t &= \sup_i \sum_{j=1}^n \|\xi_{ij}\|^2 = \sup_i \sum_{j=1}^n \|\mathbb{E}_{X+a_t\sim\mu_t}[\langle X, e_i\rangle\langle X, e_j\rangle X]\|^2 \\
&= \sup_i \sum_{j=1}^n \sum_{k=1}^n \left\langle \mathbb{E}_{X+a_t\sim\mu_t}[\langle X, e_i\rangle\langle X, e_j\rangle X], e_k\right\rangle^2 \\
&= \sup_i \sum_{j,k=1}^n \mathbb{E}_{X+a_t\sim\mu_t}[\langle X, e_i\rangle\langle X, e_j\rangle\langle X, e_k\rangle]^2 \\
&\le \sup_{\theta\in S^{n-1}} \sum_{i,j=1}^n \mathbb{E}_{X+a_t\sim\mu}[\langle X, e_i\rangle\langle X, e_j\rangle\langle X, \theta\rangle]^2.
\end{aligned}$$

We note that we cannot direct apply Lemma 4.7 at this point since the measure μ_t might not be isotropic. Hence, to be able to use the lemma, we need to normalize the covariance of μ_t. Namely, taking $X + a_t \sim \mu_t$, we define $Y = A^{-1/2}X$ which by construction is isotropic. Thus, by observing that

$$\mathbb{E}_{X+a_t\sim\mu}[\langle X, e_i\rangle\langle X, e_j\rangle\langle X, \theta\rangle]^2 \le \|A_t\|_{\text{op}}^3 \mathbb{E}_{X+a_t\sim\mu}[\langle Y, e_i\rangle\langle Y, e_j\rangle\langle Y, \theta\rangle]^2,$$

we have

$$\begin{aligned}
K_t &\le \sup_{\theta\in S^{n-1}} \sum_{i,j=1}^n \mathbb{E}_{X+a_t\sim\mu}[\langle X, e_i\rangle\langle X, e_j\rangle\langle X, \theta\rangle]^2 \\
&\le \|A_t\|_{\text{op}}^3 \sup_{\theta\in S^{n-1}} \sum_{i,j=1}^n \mathbb{E}_{X+a_t\sim\mu}[\langle Y, e_i\rangle\langle Y, e_j\rangle\langle Y, \theta\rangle]^2 \le 8C \sum_{k=1}^n \frac{1}{k(C_{\text{TS}}^n)^2}
\end{aligned} \tag{33}$$

where the last inequality follows as $\|A_t\|_{\text{op}} \le 2$ for all $t < \tau$.

At last, combining Equation (32) and (33), we arrive at

$$\text{Var}_\mu[\phi] \le 2 + \log n \left(4 + \overbrace{80C \sum_{k=1}^n \frac{1}{k}(C_{\text{TS}}^n)^{-2}}^{\Theta(\log n)} \right) = \Theta_n\left(\left(\frac{1}{C_{\text{TS}}^n} \log n \right)^2 \right)$$

implying there exists a constant $R > 0$ such that for all 1-Lipschitz ϕ, $\sqrt{\text{Var}_\mu[\phi]} \le R(C_{\text{TS}}^n)^{-1}\log n$, i.e. μ is $(R\log n)^{-1}C_{\text{TS}}^n$-concentrated and so, $C_{\text{TS}}^n \lesssim C_{\text{con}}^n \log n$ as required.

5 Log-Sobolev Inequality via Stochastic Localization

In this section, we will take a look at an application of the stochastic localization technic to prove a version of the log-Sobolev inequality for the log-concave measures. As demonstrated in the introduction, the log-Sobolev inequality is central to the concentration of measures as it allows us to directly compute concentrations of specific measures. Using stochastic localization, we will in this section bound the log-Sobolev constant for log-concave measures supported in a ball of fixed diameter. Namely, we will show Theorem 5 which states that $\rho_\mu \gtrsim D^{-1}$ for any isotropic log-concave measure μ. We take heavy inspiration from the proof as presented in Section 4 where we reduced the KLS conjecture to the thin-shell conjecture.

5.1 Entropy of bounded support

Before moving directly to proving Theorem 5, let us first attempt to gain some intuition in how the entropy of a general measure with bounded support behaves. In contrast to the variance, where the variance of a measure with bounded support is trivially bounded by the radius squared of said support, the bound for the entropy is less clear. Nonetheless, a similar result can be established for general measures with bounded support.

Lemma 5.1. For μ a measure on $\mathbb{R}^n$ supported in a ball of diameter $D \geq \epsilon$ for some fixed $\epsilon > 0$, there exists a constant C_ϵ (depending only on ϵ) such that

$$\mathrm{Ent}_\mu[\phi^2] \leq C_\epsilon D^2$$

for all smooth 1-Lipschitz functions $\phi : \mathbb{R}^n \to \mathbb{R}$.

Definition 5.1 (Sub-Gaussian measures in $\mathbb{R}$)**.** A measure μ on $\mathbb{R}$ is said to be sub-Gaussian with parameter $\nu > 0$ if

$$\psi_\mu(\lambda) = \log m_\mu(\lambda) = \log \mathbb{E}_{X\sim\mu}[e^{\lambda X}] \leq \frac{\lambda^2 \nu}{2}$$

for all $\lambda \in \mathbb{R}$. In this case, we denote $\mu \in \mathcal{G}(\nu)$.

Lemma 5.2 (Hoeffding's lemma, [Jog23])**.** For a measure μ on $\mathbb{R}$ with support of diameter D, we have $\mu \in \mathcal{G}(D^2/4)$.

For any smooth ϕ with support on D, we know that ϕ^2 has Lipschitz coefficient at most $2D\mathrm{Lip}(\phi)$. Thus, for our measure μ on $\mathbb{R}^n$ which has support on a ball of diameter D and any smooth 1-Lipschitz ϕ, we have that the push-forward measure of μ along ϕ^2, $(\phi^2)_*\mu$ is a measure on $\mathbb{R}$ with support with diameter at most $2D^2$. Consequently, Hoeffding's lemma tells us that $(\phi^2)_*\mu \in \mathcal{G}(D^4)$. With this in mind, we can now prove Lemma 5.1.

Proof of Lemma 5.1. For ease of notation, let us denote $f = \phi^2$. Consider

$$\exp\big(\mathrm{Ent}_\mu[f]\big) = \exp\big(\mathbb{E}_\mu[f \log f] - \mathbb{E}_\mu[f] \log \mathbb{E}[f]\big).$$

By considering that the function $x \mapsto -x \log x$ has its maximum at e^{-1}, we have

$$-\mathbb{E}_\mu[f] \log \mathbb{E}[f] \leq e^{-1}$$

and thusly $\exp\big(\mathrm{Ent}_\mu[f]\big) \leq e^{e^{-1}} \exp\big(\mathbb{E}_\mu[f \log f]\big)$ and we attempt to bound $\exp\big(\mathbb{E}_\mu[f \log f]\big)$. By Jensen's inequality, we have

$$\exp\big(\mathbb{E}_\mu[f \log f]\big) \leq \mathbb{E}_\mu[\exp(f \log f)] = \mathbb{E}_\mu[f e^f].$$

Now, by Taylor expansion,

$$\mathbb{E}_\mu[f e^f] = \mathbb{E}_\mu\left[f \sum_{n=0}^{\infty} \frac{f^n}{n!}\right] = \sum_{n=1}^{\infty} \frac{1}{(n-1)!}\mathbb{E}[f^n] \tag{34}$$

since the convergence is absolute.

On the other hand, by the above remark, $f_*\mu$ is a sub-Gaussian measure with parameter D^4. Thus, we can obtain a bound on the moments of μ. Indeed, since

$$\psi_{f_*\mu}(\lambda) = \log m_{f_*\mu}(\lambda) \leq \frac{\lambda^2 D^4}{2},$$

we have $m_{f_*\mu}(\lambda) \leq e^{\lambda^2 D^4/2}$ for all $\lambda \in \mathbb{R}$ where $m_{f_*\mu}$ is the moment generating function of $f_*\mu$. Hence, Taylor expanding both sides, we obtain

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \partial_\lambda^n m_{f_*\mu}(0) \leq \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{n!} D^{4n}$$

for all $\lambda \in \mathbb{R}$. Thus, we must have

$$\mathbb{E}_\mu[f^{2n}] = \partial_\lambda^{2n} m_{f_*\mu}(0) \leq D^{4n}$$

and

$$\mathbb{E}_\mu[f^{2n+1}] = \partial_\lambda^{2n+1} m_{f_*\mu}(0) \leq 0.$$

Substituting the above into Equation (34), we obtain

$$\mathbb{E}_\mu[f e^f] \leq \sum_{n=1}^{\infty} \frac{1}{(2n-1)!} D^{4n} = D^2 \sum_{n=1}^{\infty} \frac{1}{(2n-1)!} (D^2)^{2n-1} = D^2 \sinh D^2.$$

Thus,

$$\mathrm{Ent}_\mu[f] \leq \log\Big(e^{e^{-1}} \mathbb{E}_\mu[f e^f]\Big) \leq e^{-1} + \log\big(D^2\big) + \log\big(\sinh D^2\big).$$

Now, since $\log x + \log(\sinh x) \leq 2x$ and as we had assumed $D \geq \epsilon$, we have $e^{-1} \leq e^{-1}\epsilon^{-2}D^2$. Hence,

$$\mathrm{Ent}_\mu[f] \leq (2 + e^{-1}\epsilon^{-2})D^2 = C_\epsilon D^2$$

as desired. $\square$

Unfortunately, this approach cannot result in a log-Sobolev inequality as we would like to control the entropy of ϕ by $\mathbb{E}[\|\nabla\phi\|^2]$ rather than just its Lipschitz constant. Moreover, if ϕ has a high gradient in a small region, the bound resulting from the above is rather inefficient. Nonetheless, by restricting to the class of log-concave measures, Kannan, Lovász and Montenegro [LV07] were able to provide a bound for the log-Sobolev inequality of the same order.

Lemma 5.3 ([LV07])**.** For ν an isotropic log-concave measure on $\mathbb{R}^n$ supported in a ball of diameter D, we have

$$\mathrm{Ent}_\nu[\phi^2] \lesssim D^2 \mathbb{E}_\nu[\|\nabla\phi\|^2]$$

for all smooth functions $\phi : \mathbb{R}^n \to \mathbb{R}$.

5.2 Log-Sobolev inequality via stochastic localization

As promised in the introduction, we will in this section present a modified proof of Theorem 5. We attempt to analyze the log-Sobolev inequality using the same method as prescribed by Section 4.4. As it turns out, while the initial steps can be followed verbatim, this method runs into issues when trying to bound the entropy of the martingale term resulted from the localization. In particular, applying the Cauchy-Schwarz inequality no longer provides the desired integral as it had done for the variance case. Instead, by utilizing the Burkholder-Davis-Gundy inequality, we are able to express the entropy of the martingale $M_t := \int \phi^2 \mathrm{d}\mu_t$ in terms of the quadratic variation of a different martingale defined by $N_t := \int \phi^2 \log \phi^2 \mathrm{d}\mu_t$. Moreover, by computing directly, we find that N_t has quadratic variation bounded the integral of the entropies of the stochastic localization up to time t. Hence, by combining the two, we minimize the bound on the entropy by choosing to stop the stochastic localization at an appropriate time, resulting in the desired bound.

Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be smooth. Similar to before, we express the entropy of ϕ^2 with respect to μ as the entropy of the martingale terms, in particular, denoting the martingale $M_t := \int \phi^2 \mathrm{d}\mu_t$, we will show that

$$\mathrm{Ent}_\mu[\phi^2] = \mathrm{Ent}[M_t] + \mathbb{E}[\mathrm{Ent}_{\mu_t}[\phi^2]], \tag{35}$$

In fact, a more general fact is true. Namely, for any real-valued function Ψ, defining the Ψ-entropy of ϕ with respect to μ as

$$\mathrm{Ent}^\Psi_\mu[\phi] = \mathbb{E}_\mu[\Psi(\phi)] - \Psi(\mathbb{E}_\mu[\phi]),$$

we have that for ϕ satisfying $\Psi(\mathbb{E}_\mu[\phi]) = 0$,

$$\mathrm{Ent}^\Psi_\mu[\phi] = \mathrm{Ent}^\Psi[M_t] + \mathbb{E}[\mathrm{Ent}^\Psi_{\mu_t}[\phi]], \tag{36}$$

where $M_t := \int \phi \mathrm{d}\mu_t$ is a martingale. This generalizes both (35) and (24) by taking $\Psi(x) = x \log x$ and $\Psi(x) = x^2$ respectively.

The derivation is rather simple. By definition

$$\mathrm{Ent}^\Psi_{\mu_t}[\phi] = \mathbb{E}_{\mu_t}[\Psi(\phi)] - \Psi(\mathbb{E}_{\mu_t}[\phi]) = \mathbb{E}_{\mu_t}[\Psi(\phi)] - \Psi(M_t).$$

Thus, taking expectation on both sides,

$$\mathbb{E}_\mu[\Psi(\phi)] = \mathbb{E}[\mathbb{E}_{\mu_t}[\Psi(\phi)]] = \mathbb{E}[\Psi(M_t)] + \mathbb{E}[\mathrm{Ent}^\Psi_{\mu_t}(\phi)]. \tag{37}$$

However, since $\Psi(\mathbb{E}[\phi]) = 0$ by assumption, we have

$$\mathrm{Ent}^\Psi_\mu[\phi] = \mathbb{E}_\mu[\Psi(\phi)] - \Psi(\mathbb{E}_\mu[\phi]) = \mathbb{E}_\mu[\Psi(\phi)],$$

and since M_t is a martingale, we also have $\mathbb{E}[M_t] = \mathbb{E}[M_0] = \mathbb{E}_\mu[\phi]$,

$$\mathrm{Ent}^\Psi[\Psi(M_t)] = \mathbb{E}[\Psi(M_t)] - \Psi(\mathbb{E}[M_t]) = \mathbb{E}[\Psi(M_t)] - \Psi(\mathbb{E}_\mu[\phi]) = \mathbb{E}[\Psi(M_t)].$$

Finally, substituting the above two equations into (37), we recover Equation (36) as required.

With this in mind, we recall that in the reduction of the KLS conjecture, we proceeded to bound the two terms on the right hand side by using Proposition 3.3 and the Brascamp-Lieb inequality respectively. We cannot do this for general Ψ. However, in the case that $\Psi(x) = x \log x$, the term $\mathbb{E}[\mathrm{Ent}_{\mu_t}[\phi^2]]$ can be bounded using the following lemma.

Lemma 5.4 (Equation 2.17, [Led99]). Let $U : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function such that there exists some $\alpha > 0$ so that $U''(x) \geq \alpha \mathrm{id}_n$ for any $x \in \mathbb{R}^n$. Then, for all smooth functions $\phi : \mathbb{R}^n \to \mathbb{R}$, we have

$$\mathrm{Ent}_\nu[\phi^2] \leq \frac{2}{\alpha}\mathbb{E}_\nu[\|\nabla\phi\|^2]$$

where $\mathrm{d}\nu = e^{-U}\mathrm{dLeb}^n$ and U'' denotes the Hessian of U.

Recalling Equation (5), by substituting $U(x) = -z_t - \langle v_t, x\rangle + \frac{t}{2}\|x\|^2 + V(x)$ we observe that $U''(x) \geq \frac{t}{2}\mathrm{Id}_n + V''(x) \geq \frac{t}{2}\mathrm{Id}_n$ as V is convex. Thus, by the above lemma, as $\mathrm{d}\mu_t = e^{-U}\mathrm{dLeb}^n$, μ_t satisfies the log-Sobolev inequality of the form

$$\mathrm{Ent}_{\mu_t}[\phi^2] \leq 4t^{-1}\mathbb{E}_{\mu_t}[\|\nabla\phi\|^2]. \tag{38}$$

On the other hand, we cannot use the same approach to bound $\mathrm{Ent}[M_t]$. Indeed, while for Proposition 3.3, we were able to invoke a clever application of the Cauchy-Schwarz inequality in order to bound $[M]_t$ by $\mathrm{Var}_{\mu_t}[M_t]$ and the operator norm of the covariance matrix, the extra logarithmic term causes problem when we attempt to bound the entropyInstead, we shall bound $\mathrm{Ent}[M_t]$ by an application of the Burkholder-Davis-Gundy inequality.

Lemma 5.5 (Burkholder-Davis-Gundy, [Low10]). For all $1 \leq p < \infty$, there exists some constant C_p such that for any martingale $(X_t)_{t\geq 0}$, we have

$$\mathbb{E}\left[\left(\sup_{s\leq t} X_s\right)^p\right] \leq C_p\mathbb{E}\left[[X]_t^{p/2}\right]$$

for all $t \geq 0$.

With the above lemma in mind, denoting $\psi = \phi^2\log\phi^2$ and the martingale $N_t = \int \psi \mathrm{d}\mu_t$, we have

$$\begin{aligned}
\mathrm{Ent}[M_t] &= \mathbb{E}[M_t \log M_t] - \mathbb{E}[M_t]\overbrace{\log \mathbb{E}[M_t]}^{=0} \\
&= \mathbb{E}[M_t \log M_t] = \mathbb{E}[\mathbb{E}_{\mu_t}[\phi^2]\log\mathbb{E}_{\mu_t}[\phi^2]] \\
&\leq \mathbb{E}[\mathbb{E}_{\mu_t}[\phi^2\log\phi^2]] = \mathbb{E}[N_t] \\
&\leq C_1\mathbb{E}\left[[N]_t^{1/2}\right].
\end{aligned} \tag{39}$$

where the first inequality follows by Jensen's inequality. Now, similar to the variance case, we focus on bounding $[N]_t$. Similar to the proof of Proposition 3.3, introducing the notation $U_t := \int \psi(x)(x-a_t)\mu_t(\mathrm{d}x)$, we obtain

$$\begin{aligned}
\mathrm{d}N_t &= d\int \psi(x)F_t(x)\mu(\mathrm{d}x) = \int \psi(x)\langle x-a_t, \mathrm{d}W_t\rangle\mu_t(\mathrm{d}x) \\
&= \left\langle \int \psi(x)(x-a_t)\mu_t(\mathrm{d}x), \mathrm{d}W_t\right\rangle = \langle U_t, \mathrm{d}W_t\rangle.
\end{aligned}$$

Hence, as $\int (x-a_t)\mu_t(\mathrm{d}x) = 0$, we have

$$\begin{aligned}
\mathrm{d}[N]_t = \|U_t\|^2\mathrm{d}t &= \left\|\int \psi(x)(x-a_t)\mu_t(\mathrm{d}x)\right\|^2 \mathrm{d}t \\
&= \left\|\int (\psi(x) - M_t\log M_t)(x-a_t)\mu_t(\mathrm{d}x)\right\|^2 \mathrm{d}t.
\end{aligned}$$

Now, by recalling that any linear tilt localization satisfies $\mu_t \ll \mu$ and moreover μ support in a ball of diameter D, μ_t must therefore also has support in a ball of diameter D. Thus, $\|x - a_t\|_{L^\infty(\mu_t)} \le D$ for any $t \ge 0$. Then, we have

$$\begin{aligned}
\mathrm{d}[N]_t &= \left\| \int (\psi(x) - M_t \log M_t)(x - a_t)\mu_t(\mathrm{d}x) \right\|^2 \mathrm{d}t \\
&\le D^2 \left\| \int (\psi(x) - M_t \log M_t)\mu_t(\mathrm{d}x) \right\|^2 \mathrm{d}t \\
&= D^2 \left| \mathbb{E}_{\mu_t}[\phi^2 \log \phi^2] - \mathbb{E}_{\mu_t}[\phi^2] \log \mathbb{E}_{\mu_t}[\phi^2] \right|^2 \mathrm{d}t \\
&= D^2 \mathrm{Ent}_{\mu_t}[\phi^2]^2 \mathrm{d}t.
\end{aligned}$$

Combining this with Equation (39), we obtain

$$\mathrm{Ent}[M_t] \lesssim \mathbb{E}\left[\int_0^t \mathrm{d}[N]_s^{1/2} \right] \le \mathbb{E}\left[\int_0^t D\mathrm{Ent}_{\mu_s}[\phi^2]\mathrm{d}s \right]. \tag{40}$$

At this point, as μ_t is only supported in a ball of diameter D, we can utilize Lemma 5.3 for bounding the log-Sobolev constant. Hence, combining this with Lemma 5.4, we have that $\mathrm{Ent}_{\mu_t}[\phi^2] \lesssim (t^{-1} \wedge D)\mathbb{E}_{\mu_t}[\|\nabla \phi\|^2]$. Thus, by Fubini's theorem, Equation (40) becomes

$$\begin{aligned}
\mathrm{Ent}[M_t] &\lesssim \mathbb{E}\left[D \int_0^t (s^{-1} \wedge D^2)\mathbb{E}_{\mu_s}[\|\nabla \phi\|^2]\mathrm{d}s \right] \\
&= D \int_0^t (s^{-1} \wedge D^2)\mathbb{E}_{\mu}[\|\nabla \phi\|^2]\mathrm{d}s \\
&= D\mathbb{E}_{\mu}[\|\nabla \phi\|^2] \left(1 + \int_{D^{-2}}^t s^{-1}\mathrm{d}s \right) \\
&= D\mathbb{E}_{\mu}[\|\nabla \phi\|^2](1 + \log t + 2\log D)
\end{aligned}$$

for all $t > D^{-2}$.

Combining this with Equation (35), we obtain

$$\begin{aligned}
\mathrm{Ent}_{\mu}[\phi^2] &\le \mathrm{Ent}[M_t] + \mathbb{E}[\mathrm{Ent}_{\mu_t}[\phi^2]] \\
&\lesssim \mathbb{E}_{\mu}[\|\nabla \phi\|^2](D(1 + \log t + \log D) + 4t^{-1})
\end{aligned}$$

for all $t > D^{-2}$. Minimizing this equation in t, we find that it is minimized when $t = 4D^{-1}$ resulting in

$$\mathrm{Ent}_{\mu}[\phi^2] \lesssim (2 + \log 4)D\mathbb{E}_{\mu}[\|\nabla \phi\|^2]$$

whenever $4D^{-1} \ge D^{-2}$ which holds if and only if $D \ge 1/4$. Thus,

$$\mathrm{Ent}_{\mu}[\phi^2] \lesssim D\mathbb{E}_{\mu}[\|\nabla \phi\|^2]$$

where for $D \le 1$, $D^{-2} \ge D^{-1}$ and so the conclusion remains to hold by Lemma 5.3. Hence, we bounded the log-Sobolev constant with $\rho_\mu \gtrsim D^{-1}$ recovering the result of Theorem 5.

References

[ALOG20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model, 2020.

[BL76] Herm Brascamp and Elliott 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. *Journal of Functional Analysis*, 22(4):366–389, 1976.

[Bog98] Vladimir Bogachev. *Gaussian measures*. American Mathematical Society, 1998.

[Bou86] J. Bourgain. On high dimensional maximal functions associated to convex bodies. *American Journal of Mathematics*, 108(6):1467–1476, 1986.

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains, 2022.

[Che20] Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the kls conjecture, 2020.

[Eld13] Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, mar 2013.

[Eld18] Ronen Eldan. Lecture notes - from stochastic calculus to geometric inequalities, 2018.

[Gro75] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[Jog23] Varun Jog. Concentration inequalities, 2023.

[Kla06] Boáz Klartag. Power-law estimates for the central limit theorem for convex sets, 2006.

[Led99] Michel Ledoux. Concentration of measure and logarithmic sobolev inequalities. In Jacques Azéma, Michel Émery, Michel Ledoux, and Marc Yor, editors, *Séminaire de Probabilités XXXIII*, pages 120–216, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[Low10] George Lowther. The burkholder-davis-gundy inequality, 2010.

[LPW17] David Asher Levin, Yuval Peres, and Elizabeth Wilmer. *Markov chains and mixing times*. American Mathematical Society, second edition. edition, 2017.

[LV07] László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307358, may 2007.

[LV16] Yin Tat Lee and Santosh S. Vempala. Eldan's stochastic localization and the kls conjecture: Isoperimetry, concentration and mixing, 2016.

[LV18] Yin Tat Lee and Santosh S. Vempala. The kannan-lovász-simonovits conjecture, 2018.

[Mil08] Emanuel Milman. Uniform tail-decay of lipschitz functions implies cheeger's isoperimetric inequality under convexity assumptions. *Comptes Rendus Mathematique*, 346(17):989–994, 2008.

[Mil18] Emanuel Milman. On the role of convexity in isoperimetry, spectral-gap and concentration. 2018.

[Øks03] Bernt Øksendal. *Stochastic Differential Equations*, pages 65–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[Yan19] Junchi Yang. Herbst's argument, and lsi to poincare. 2019.

A Proof of Herbst's Argument

We will in this appendix provide a short proof adapted from [Yan19] of Herbst's argument.

Theorem 10 (Herbst's argument)**.** If μ satisfy the log-Sobolev inequality with log-Sobolev constant ρ_μ, then for all ϕ with uniformly bounded gradient $\|\nabla\phi\| \leq K$, we have

$$\psi^{\mu}_{\phi-\mathbb{E}_\mu[\phi]}(\lambda) \leq \frac{K^2\lambda^2}{2\rho_\mu}$$

where $\psi^{\mu}_{\phi-\mathbb{E}_\mu\phi}$ is the logarithmic moment generating function of $\phi - \mathbb{E}_\mu[\phi]$ with respect to μ.

To prove Herbst's argument, we will first need the following lemma which allows us to ignore the normalization condition that $\int \phi^2 \mathrm{d}\mu = 1$.

Lemma A.1. If μ satisfy the log-Sobolev inequality with log-Sobolev constant ρ_μ, then for all smooth $\phi : \mathbb{R}^n \to \mathbb{R}$ (in particular, we do not require $\int \phi^2 = 1$), we have

$$\frac{\rho_\mu}{2}\mathrm{Ent}_\mu[\phi^2] \leq \mathbb{E}[\|\nabla\phi\|^2].$$

Proof. Denoting $\psi = \phi/\sqrt{\mathbb{E}_\mu[\phi^2]}$, (noting that $\mathbb{E}_\mu[\phi^2] > 0$ except for $\phi = 0$ almost everywhere for which the inequality is trivial) we have by the log-Sobolev inequality that

$$\frac{\rho_\mu}{2}\mathrm{Ent}_\mu[\psi^2] \leq \mathbb{E}_\mu[\|\nabla\psi\|^2]. \tag{41}$$

On the other hand, by definition

$$\mathrm{Ent}_\mu[\psi^2] = \mathbb{E}_\mu\left[\frac{\phi^2}{\mathbb{E}_\mu[\phi^2]}\log\left(\frac{\phi^2}{\mathbb{E}_\mu[\phi^2]}\right)\right] = \frac{1}{\mathbb{E}_\mu[\phi^2]}\mathrm{Ent}_\mu[\phi^2].$$

Hence, by substituting the above into Equation 41, it follows that

$$\frac{\rho_\mu}{2}\mathrm{Ent}_\mu[\phi^2] \leq \mathbb{E}_\mu[\phi^2]\mathbb{E}_\mu\left[\left\|\nabla\frac{\phi}{\sqrt{\mathbb{E}_\mu[\phi^2]}}\right\|^2\right] = \mathbb{E}_\mu[\|\nabla\phi\|^2]$$

as required. $\square$

Proof of Herbst's argument. Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a smooth function with $\|\nabla\phi\| \leq K$. Moreover, without loss of generality we will assume ϕ satisfies $\mathbb{E}_\mu[\phi] = 0$.

For all $\theta \in \mathbb{R}$, let us denote $\xi_\theta = \exp\left(\frac{\theta}{2}\phi\right)$. Then, applying the log-Sobolev inequality to ξ_θ we have,

$$\frac{\rho_\mu}{2}\mathrm{Ent}_\mu[\xi_\theta^2] \leq \mathbb{E}_\mu[\|\nabla\xi_\theta\|^2].$$

Then, denoting the moment generating function $m(\theta) = \mathbb{E}_\mu[\exp(\theta\phi)] = \mathbb{E}_\mu[\xi_\theta^2]$, we observe

$$\theta m'(\theta) = \mathbb{E}_\mu[\theta\phi\exp(\theta\phi)]$$

and so, the entropy of ξ_θ^2 can be equivalently be written as

$$\begin{aligned}\mathrm{Ent}_\mu[\xi_\theta^2] &= \mathbb{E}_\mu[\xi_\theta^2 \log \xi_\theta^2] - \mathbb{E}_\mu[\xi_\theta^2] \log\big(\mathbb{E}_\mu[\xi_\theta^2]\big) \\ &= \theta m'(\theta) - m(\theta) \log m(\theta).\end{aligned}$$

On the other hand, we observe

$$\mathbb{E}_\mu[\|\nabla \xi_\theta\|^2] = \mathbb{E}_\mu\left[\sum_{i=1}^n \left(\frac{\theta}{2}\xi_\theta \partial_i \phi\right)^2\right] = \left(\frac{\theta}{2}\right)^2 \mathbb{E}_\mu[\xi_\theta^2 \|\nabla \phi\|^2] \leq \left(\frac{K\theta}{2}\right)^2 m(\theta)$$

where the last inequality follows as $\|\nabla \phi\| \leq K$. Hence, combining the above two equations into the log-Sobolev inequality provides

$$\frac{\rho_\mu}{2}(\theta m'(\theta) - m(\theta) \log m(\theta)) \leq \left(\frac{K\theta}{2}\right)^2 m(\theta).$$

Rearranging, we obtain

$$\left(\frac{1}{\theta} \log m(\theta)\right)' \leq \frac{K^2}{2\rho_\mu}$$

where the derivative is taken with respect to θ. Thus, since as a consequence of l'Hôpital's rule,

$$\lim_{\theta \to 0} \frac{1}{\theta} \log m(\theta) = \frac{m'(0)}{m(0)} = \mathbb{E}_\mu[\phi] = 0,$$

it follows by integrating both sides of the above inequality from 0 to λ that

$$\log m(\lambda) \leq \frac{K^2 \lambda^2}{2\rho_\mu}$$

as claimed. $\square$