Stochastic Localization and its Applications to Markov Mixing and the KLS Conjecture

Kexing Ying

March 18, 2023

Contents

1 Introduction

Sampling from a given distribution is a fundamental problem arising naturally in many fields such as bandit theory, machine learning, Bayesian statistics, etc. Among these problems, sampling from the class of distributions known as the log-concave distributions are of particular interest. These distributions characterizes the uniform distributions on convex bodies and is related to the volumes of said bodies. Beyond this, sampling from the log-concave measures is useful for optimization, e.g. minimizing convex functions.

Sampling from these distributions has been studied for a long time and there are many algorithms devised to do so.

1.1 Structure of this essay

We will in this essay examine three related applications of the stochastic localization scheme. In particular, we will consider its application to providing Markov mixing bounds, the relation of the KLS and thin-shell conjectures, and its application to proving a log-Sobolev inequality.

Markov mixing bounds:

KLS and thin-shell conjectures:

Log-Sobolev inequality: The log-Sobolev inequality is a class of inequalities fundamental in the theory of concentration of measures.

2 Stochastic Localization Scheme

In this section we introduce the notion of stochastic localization schemes. To gain an intuition for these objects, we also present several examples which is be studied further in subsequent sections.

More intro

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 $(\Omega, \mathscr{F}, \mathbb{P})$ 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).

Where $\mathscr{M}(\mathscr{X})$ is equipped with the σ-algebra generated by maps of the form

$$\pi_A : \mathscr{M}(\mathscr{X}) \mapsto \mathbb{R}_{\geq 0} \cup \{\infty\} : \mu \mapsto \mu(A)$$

for all $A \in \Sigma$. Equivalently, this is the Borel σ-algebra on $\mathscr{M}(\mathscr{X})$ using the topology induced by the total variation norm.

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 $t \to \infty$.

Definition 2.3 (Stochastic localization scheme, [CE22])**.** Denoting $\mathscr{L}(\mu)$ the set of all stochastic localization processes of the measure μ, a stochastic localization scheme is a map

$$\Phi : \mathscr{M}(\mathscr{X}) \to \coprod_{\mu \in \mathscr{M}(\mathscr{X})} \mathscr{L}(\mu)$$

such that $\Phi(\mu) \in \mathscr{L}(\mu)$ for all $\mu \in \mathscr{M}(\mathscr{X})$.

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 $t \geq 0$.
- taking $X \sim \mu$ such that $\mu_t \to \delta_X$ almost everywhere as $t \to \infty$ (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 $X : \omega \mapsto x_\omega$, it suffices 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. □

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 shall examine the property in section 3.3, though we will construct the scheme 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 $[n] = \{1, \cdots, n\}$ independent of X. Then, the coordinate by coordinate stochastic localization of μ is the process $(\mu_k)_k$ such that for all $x \in \{-1, 1\}^n$,

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

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

$(\mu_k)_k$ is indeed a stochastic localization of μ. It is clear that (L0) and (L1) are satisfied. By construction of $(\mu_k)_k$, denoting $\mathscr{F}_k := \sigma(X_{Y_1}, \cdots, X_{Y_{n\wedge k}})$, 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 $\{Y_1, \cdots, Y_n\}$ 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 $\langle X, Y_1\rangle, \cdots, \langle X, Y_{n\wedge k}\rangle$.

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 these 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 2.4 (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 2.5 (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

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

   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 (1) 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]):

$$F_0 = 1, \mathrm{d}F_t(x) = \langle x - \bar{\mu}(F_t), \mathrm{d}W_t\rangle F_t(x), \tag{2}$$

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 $x \in \mathbb{R}^n$,

$$\mathrm{d}\log F_t(x) = \frac{\mathrm{d}F_t(x)}{F_t(x)} - \frac{\mathrm{d}[F(x)]_t}{2F_t(x)^2} = \langle x - a_t, \mathrm{d}W_t\rangle - \frac{1}{2}\|x - a_t\|^2 \mathrm{d}t \tag{3}$$

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 of the form

$$F_t(x) = e^{z_t + \langle x, v_t\rangle - \frac{t}{2}\|x\|^2}, \tag{4}$$

for given Itô processes $(z_t), (v_t)$.
With this formulation of $F_t(x)$ in mind, it follows 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 if μ 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.

Definition 2.6 (Log-concave measure)**.** A measure μ on $\mathbb{R}^n$ is said to log-concave if it is of the form $\mathrm{d}\mu = \exp(-H)\mathrm{dLeb}^n$ for some convex function $H : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$

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:

$$A_t := \mathrm{Cov}[\mu_t] = \int (x - a_t) \otimes (x - a_t) \mu_t(\mathrm{d}x), \tag{5}$$

where $\otimes$ denotes the Kronecker product. In particular, we will show $(A_t)_{ij} \to 0$ for all $i, j \in \{1, \cdots, n\}$ 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 inequality, [BL76])**.** Given $V : \mathbb{R}^n \to \mathbb{R}$ convex and $K > 0$, if ν is an isotropic (i.e. $\mathbb{E}_{X\sim\nu}[X] = 0$ and $\mathrm{Cov}_{X\sim\nu}(X) = \mathrm{id}_n$) 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 Poincaré inequality. In particular, 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 (4) and defining $\pi_i(x) := x_i$, we have by the Cauchy-Schwarz inequality

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

Again, using equation (4), we note that any realizations of $(F_t(x))$ is eventually decreasing in t for all $x \neq 0$, implying

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

Thus, by taking $t \to \infty$ we have $(A_t)_{ij} \to 0$ for all $i, j \in \{1, \cdots, n\}$ 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

$$M_t \xrightarrow{\text{a.e.}} \phi(a_\infty) \sim \phi_* \mu \tag{6}$$

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 (1) can be rewritten as

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

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

Definition 2.7 (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

$$\mathrm{d}\mu_{k+1} = (1 + \langle x - \bar{\mu}_k, Z_k \rangle)\mathrm{d}\mu_k \tag{7}$$

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

$$Z_k := e_{Y_k} \cdot \begin{cases} \frac{1}{1+(\bar{\mu}_k)_{Y_k}} & \text{with probability } \frac{1+(\bar{\mu}_k)_{Y_k}}{2} \\ \frac{-1}{1-(\bar{\mu}_k)_{Y_k}} & \text{with probability } \frac{1-(\bar{\mu}_k)_{Y_k}}{2} \end{cases} \tag{8}$$

where $e_1, \cdots, e_n$ 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 a method to compute a bound on the mixing time for a special class of Markov processes; namely 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, recovering the main theorem on spectral independence as presented by Anari, Liu and Oveis Gharan in [ALOG20].

3.1 Mixing bounds

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. 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 (X_n) instead. It is in general not difficult to come up with such Markov processes, although it is often difficult to show 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 3.1 (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

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

Furthermore, we denote

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

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 3.2 (Spectral gap, [Lev17])**.** The spectral gap of a Markov kernel K is defined to be

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

It is not difficult to show that

$$\mathfrak{gap}(K) = \inf_{\substack{\phi:\mathscr{X}\to\mathbb{R} \\ \int \phi \mathrm{d}\mu = 0}} 1 - \frac{\int \phi K \phi \mathrm{d}\mu}{\int \phi^2 \mathrm{d}\mu} = \inf_{\phi:\mathscr{X}\to\mathbb{R}} \frac{1}{2\mathrm{Var}_\mu[\phi]} \int (\phi(x) - \phi(y))^2 K(x, \mathrm{d}y)\mu(\mathrm{d}x), \tag{9}$$

where μ is the stationary measure of K and $K\phi = \int \phi(y)K(\cdot, \mathrm{d}y)$. We will take equation (9) to be the defining property of the spectral gap in the case K is defined on a general state space.

Theorem 2 ([Lev17])**.** 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

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

We remark that this inequality is only meaningful whenever $\mu_{\min} > 0$ and thus, this theorem is only meaningful for Markov chains on finite state space (and we can replace inf with min).

3.1.1 Ising model and Glauber dynamics

Definition 3.3 (Ising model)**.** 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.

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 β in which a larger β means that neighboring vertices are more likely to align.

As illustrated by theorem 1, in order to sample from the Ising model, we can construct a Markov chain which has the Ising model as its stationary distribution. One such Markov chain is known as the Glauber dynamics.

Definition 3.4 (Glauber dynamics)**.** 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 $\sigma^1, \sigma^2 \in \{-1,1\}^n$.

Parsing this definition, we see that the Glauber dynamics is the Markov chain such that, starting at a configuration $\sigma_1 \in \{-1,1\}^n$, 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 μ. One may easily compute that the Glauber dynamics of the Ising model $\mu_{\beta,h}$ indeed has $\mu_{\beta,h}$ as its stationary distribution.

Recently, a framework known as spectral independence has been developed by Anari, Liu and Oveis Gharan in [ALOG20] which provides a bound for the spectral gap and consequently a mixing bound by for the Glauber dynamics (namely, theorem 3).

Definition 3.5 (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 3.6 (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 $(\eta_0, \cdots, \eta_{n-2})$-spectrally independent if defining μ_k to be the law of X conditioned on $X_{i_1}, \cdots, X_{i_k}$ for some $X \sim \mu$ and any $\{i_1, \cdots, i_k\} \subseteq \{1, \cdots, n\}$, μ_k is η_k-spectrally independent for each $k = 0, \cdots, n-2$.

Theorem 3 ([ALOG20]). If μ is a measure on $\{-1,1\}^n$ which is $(\eta_0, \cdots, \eta_{n-2})$-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 an equivalent bound by using the theories established above. In particular, assuming the same assumptions as theorem 3, we 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 previously, 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 3.7 (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 (almost everywhere) with respect to μ for all t. 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 $x \in \mathscr{X}, A \in \Sigma$.

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 is indeed 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 as 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

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

Similarly, by the same calculation,

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

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. $\square$

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) (where we take $\phi(x,y) = \phi(x)\phi(y)$), we have

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

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

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

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

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

Hence, recalling the equivalent form of the spectral gap as described by (9), the result follows by taking infimum on both sides. □

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 $\epsilon>0, k=0,\cdots,m-1$ 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 3.8 (Discrete time approximate conservation of variance (ACV), [CE22])**.** A prelocalization process $(\mu_k)_k$ is said satisfy conserve (κ_k)-variance up to time m if for any integrable function $\phi:\mathscr{X}\to\mathbb{R}$, $0\leq k<m$,

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

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

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

$$
\begin{aligned}
\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_k] &= \mathbb{E}\left[\int \phi^2 \mathrm{d}\mu_{k+1} - \left(\int \phi \mathrm{d}\mu_{k+1}\right)^2 \,\middle|\, \mu_k\right] \\
&= \int \phi^2 \mathrm{d}\mu_k - \mathbb{E}\left[\left(\int \phi(x)(1+\langle x-\bar{\mu}_k, Z_k\rangle)\mu_k(\mathrm{d}x)\right)^2 \,\middle|\, \mu_k\right] \\
&= \int \phi^2 \mathrm{d}\mu_k - \left(\int \phi \mathrm{d}\mu_k\right)^2 - \mathbb{E}\left[\left(\int \phi(x)\langle x-\bar{\mu}_k, Z_k\rangle\mu_k(\mathrm{d}x)\right)^2 \,\middle|\, \mu_k\right] \\
&= \mathrm{Var}_{\mu_k}[\phi] - \mathrm{Var}\left[\int \phi(x)\langle x-\bar{\mu}_k, Z_k\rangle\mu_k(\mathrm{d}x) \,\middle|\, \mu_k\right]
\end{aligned} \tag{13}
$$

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 (13) 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

$$\mathbb{E}[\mathrm{Var}_{\mu_{k+1}}[\phi] \mid \mu_k] \ge (1 - \|C_k^{1/2}\mathrm{Cov}(\mu_k)C_k^{1/2}\|_{\mathrm{op}})\mathrm{Var}_{\mu_k}[\phi] \tag{14}$$

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 $\mathrm{d}\mathrm{Var}[M_t]$.

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

$$\mathrm{d}\mathrm{Var}[M_t] \leq \mathrm{Var}[M_t]\mathbb{E}[\|A_t\|_{\mathrm{op}}]\mathrm{d}t,$$

where A_t is the covariance matrix of μ_t and $M_t := \int \phi \, \mathrm{d}\mu_t$.

Unify covariance notation.

To prove 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, as ϕ is isotropic, $M_0 = \int \phi \, \mathrm{d}\mu = 0$ and so, $\mathrm{Var}[M_t] = \mathbb{E}[M]_t$. As a result of this, we will now analyze $\mathrm{d}[M]_t$.

Proof. With the above remarks in mind, by taking expectation, it is sufficient to bound $\mathrm{d}[M]_t$ instead. Observe that

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

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

$$\mathrm{d}[M]_t = \|U_t\|^2 \, \mathrm{d}t. \tag{15}$$

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 observe,

$$\begin{aligned} \mathrm{d}[M]_t = \langle U_t, \theta \rangle^2 \, \mathrm{d}t &= \left\langle \int \left(\phi(x) - \int \phi \, \mathrm{d}\mu_t \right)(x - a_t)\mu_t(\mathrm{d}x), \theta \right\rangle^2 \mathrm{d}t \\ &= \left(\int \left(\phi(x) - \int \phi \, \mathrm{d}\mu_t \right) \langle x - a_t, \theta \rangle \mu_t(\mathrm{d}x) \right)^2 \mathrm{d}t \\ &\leq \mathrm{Var}_{\mu_t}[\phi] \left(\int \langle \theta, (x - a_t)^{\otimes 2} \theta \rangle \mu_t(\mathrm{d}x) \right) \mathrm{d}t \\ &= \mathrm{Var}_{\mu_t}[\phi] \langle \theta, A_t \theta \rangle \mathrm{d}t \leq \mathrm{Var}_{\mu_t}[\phi] \|A_t\|_{\mathrm{op}} \mathrm{d}t. \end{aligned} \tag{16}$$

where again, the inequality follows by the Cauchy-Schwarz inequality. Thus, by taking expectation on both sides of the inequality, we obtain the required inequality. □

3.3 Glauber dynamics as an associated Markov process

As alluded to in section 2, 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

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

Then, substituting in the following equalities:

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

and

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

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 (14) that the spectral gap of the Glauber dynamics is bounded below by

$$\mathfrak{gap}(K) \geq \prod_{k=0}^{n-2}(1 - \|C_k^{1/2}\mathrm{Cov}(\mu_k)C_k^{1/2}\|_{\mathrm{op}}), \tag{17}$$

where $C_k := \mathrm{Cov}[Z_k \mid \mu_k]$ with Z_k defined by equation (8). 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

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

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

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

we have $C_k^{1/2}\mathrm{Cov}(\mu_k)C_k^{1/2} = \frac{1}{n-k}\mathrm{Cor}(\mu_k)$ and equation (17) becomes

$$\mathfrak{gap}(K) \geq \prod_{k=0}^{n-2}\left(1 - \frac{\|\mathrm{Cor}(\mu_k)\|_{\mathrm{op}}}{n-k}\right). \tag{18}$$

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 $(\eta_0,\cdots,\eta_{n-2})$-spectrally independent, we have $\rho(\Psi(\mu_k)) \leq \eta_k$ for all $k = 0,\cdots,n-2$ and so, by equation (17), we have

$$\mathfrak{gap}(K) \geq \prod_{k=0}^{n-2}\left(1 - \frac{\eta_k + 1}{n-k}\right). \tag{19}$$

This bound and theorem 3 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

In this section, we present the original context and motivation for the construction of stochastic localizations: the KLS conjecture. The KLS conjecture is a conjecture which roughly states that all log-concave measures on $\mathbb{R}^n$ are concentrated in a way similar to a Gaussian measure. To tackle this problem, Eldan in [Eld13] introduced stochastic localizations in order to reduce the KLS conjecture to a seemingly weaker conjecture (up to a logarithmic constant): the thin-Shell conjecture. More recently, stochastic localization was applied by Chen in [Che20] to provide an almost constant lower bound for the KLS conjecture. We will now 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 [Eld18].

4.1 Concentration

Let us quickly introduce some preliminary definitions required to state the aforementioned conjectures.

Definition 4.1 (Concentration, [Eld18])**.** Let μ be a measure on $\mathbb{R}^n$, then μ is said to be C-(inversely)-concentrated if for all 1-Lipschitz function $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Var}_\mu[\phi] = \mathrm{Var}_{X\sim\mu}[\phi(X)] \leq C^2. \tag{20}$$

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

Heuristically, the concentration is a measure of the variance of μ in each direction and in particular equals the variance in the 1-dimensional case. 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 ϕ : $\mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Var}[\phi(X)] \leq K^2\mathrm{Var}[X].$$

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

$$\begin{aligned}
2\mathrm{Var}[\phi(X)] &= \mathrm{Var}[\phi(X) - \phi(X')] && \text{(i.i.d.)}\\
&= \mathbb{E}[(\phi(X) - \phi(X'))^2] - \mathbb{E}[\phi(X) - \phi(X')]^2\\
&= \mathbb{E}[(\phi(X) - \phi(X'))^2] && \text{(identically distributed)}\\
&\leq K^2\mathbb{E}[\|X - X'\|^2] && \text{(as } \phi \text{ is } K\text{-Lipschitz)}\\
&= K^2\mathbb{E}[X^TX + X'^TX' - X^TX' - X'^TX]\\
&= 2K^2\mathrm{Var}[X] - 2K^2\mathrm{Cov}(X, X') = 2K^2\mathrm{Var}[X]. && \text{(independence)}
\end{aligned}$$

implying $\mathrm{Var}[\phi(X)] \leq K^2\mathrm{Var}[X]$ 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

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

as required. $\square$

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

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 definitions in a specific sense.

Definition 4.2 (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

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

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

Definition 4.3 (First-moment concentration, [Mil18]). Again, given 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 4 (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, $D^{\mu}_{\exp} \simeq D^{\mu}_{\text{FM}}$ 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 $D^{\mu}_{\exp} \leq \sqrt{2}(C^{\mu}_{\text{con}})^{-1}$ and $(2C^{\mu}_{\text{con}})^{-1} \leq D^{\mu}_{\text{FM}}$. Hence, if μ is log-concave, $D^{\mu}_{\exp} \simeq D^{\mu}_{\text{FM}} \simeq (C^{\mu}_{\text{con}})^{-1}$.

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

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

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

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

implying μ is first-moment concentrated with respect to the constant $(2C)^{-1}$.

On the other hand, if μ is exponential concentration 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 $\sqrt{2}D^{-1}$-concentrated. □

4.2 Example: concentration of the Gaussian

Given a class of measures, it is in general *not* true that the concentration coefficient of measures of said class is dimension invariant. However, this turns out to be the case for the Gaussian measures.

Theorem 5 (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 $\phi : \mathbb{R}^n \to \mathbb{R}$, $\mathrm{Var}_{\gamma^n}[\phi] \leq C^2$.

This fact motivate the KLS conjecture which hypothesized that this invariance holds for a larger class of measures known as the log-concave measures. We will for completeness give a brief proof (based on theorem 1.7.1 in [Bog98]) of the above theorem taking $C = \pi/2$ (although one can further show Gaussian measures are 1-concentrated).

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 $f_x^*\gamma^n \sim \mathcal{N}(0, \|x\|^2)$ as f_x^* is linear. 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

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

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

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

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 5 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 The KLS and thin-shell conjecture

As alluded to in the previous section, the KLS conjecture suggests that any log-concave measure on $\mathbb{R}^n$ admits concentration roughly similar to that of the Gaussian measure. However, unlike the Gaussian, as the concentration of measures is not invariant under linear functions, it is clear that the KLS conjecture would not hold without a suitable normalization. This leads us to the following formulation of the KLS conjecture.

Conjecture 1 (Kannan-Lovász-Simonovitz, [Eld18]). Denoting $\mathcal{M}^n_{\text{con}}$ the set of all log-concave probability measures μ on $\mathbb{R}^n$ satisfying $\text{Var}_\mu[T] \leq 1$ for all 1-Lipschitz linear maps $T : \mathbb{R}^n \to \mathbb{R}$, there exists a *universal* constant C (i.e. does not depend on any parameter and in particular is independent of the dimension n) such that for all $\mu \in \mathcal{M}^n_{\text{con}}$, μ is C-concentrated.

Assuming the normalization condition, 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 $\text{Var}_{X\sim\mu}[X_i] \leq 1$. Thus,

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

Hence, by proposition 4.1, we have that $\text{Var}_\mu[\phi] \leq n$ for all 1-Lipschitz $\phi : \mathbb{R}^n \to \mathbb{R}$ implying μ is $\sqrt{n}$-concentrated.

Conjecture 2 (Thin-shell, [Eld13]). Taking $\mathcal{M}^n_{\text{con}}$ as above, there exists a universal constant C such that for all $\mu \in \mathcal{M}^n_{\text{con}}$, we have

$$\sqrt{\text{Var}_\mu[\|\cdot\|]} \leq C.$$

Since the norm function is 1-Lipschitz, it is clear that the thin-shell conjecture is weaker than that of the KLS conjecture. On the other hand, as we shall describe in the following subsections, as a consequence of the theory of stochastic localization, Eldan [Eld13] provides a reduction of the KLS conjecture to the thin-shell conjecture up to logarithmic factors.

Theorem 6 (Eldan, [Eld13]). Denoting $\mathcal{M}^n_{\text{con}}$ as above, we define

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

and

$$C^n_{\text{TS}} := \inf\left\{C \mid \forall \mu \in \mathcal{M}^n_{\text{con}},\ \sqrt{\text{Var}_\mu[\|\cdot\|]} \leq C\right\} = \sup_{\mu \in \mathcal{M}^n_{\text{con}}} \sqrt{\text{Var}_\mu[\|\cdot\|]},$$

we have,

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

The stochastic localization scheme has been wildly successful in making progress towards the KLS conjecture. Modifying the original arguments by Eldan, Lee and Vempala [LV16] obtained the bound $C^n_{\text{con}} \lesssim n^{-1/4}$. Further modifying their arguments, a recent breakthrough by Chen [Che20] improves the bound providing the following theorem.

Theorem 7 (Chen, [Che20]). $\log C^n_{\text{con}} \lesssim \sqrt{\log n \log\log n}$ and so $C^n_{\text{con}} = n^{-o(1)}$.

4.3.1 Equivalent formulation of the KLS conjecture

While we have formulated the KLS conjecture using the language of concentration, the conjecture itself was originally formulated as an isoperimetric problem. For completeness of this exposition, we shall briefly present these equivalent formulations here.

The isoperimetric problem is the problem in finding the set of unit volume with minimum surface area. In the case of the $\mathbb{R}^n$ equipped with the Lebesgue measure, we have known since the ancient Greeks [Bl05] that the solution is the unit ball. With this in mind, it is natural for us to generalize the problem for arbitrary measures.

Definition 4.4 (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,

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

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

The isoperimetric problem for the measure μ then becomes the problem of finding the set A satisfying $\mu(A) = 1$ with minimum $\mu^+(\partial A)$.

Definition 4.5 (Cheeger's inequality, [Mil08])**.** Given a measure μ on $\mathbb{R}^n$, we say μ satisfy Cheeger's inequality if there exists some D such that for all A,

$$\mu(A) \wedge \mu(A^c) \leq D\mu^+(A).$$

We call the largest such D the inverse Cheeger's constant (or the inverse isoperimetric constant) and denote it by D^μ_{C}.

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

Conjecture 3 (KLS, [Eld13])**.** Denoting $\mathscr{M}^n_{\mathrm{iso}}$ 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 $\mu \in \mathscr{M}^n_{\mathrm{iso}}$, μ satisfy the Cheeger's inequality with constant D.

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

Theorem 8 (Milman, [Mil08])**.** For all log-concave measure μ on $\mathbb{R}^n$, the following are equivalent

- μ has exponential concentration with constant D^μ_{exp}.
- μ has first-moment concentration with constant D^μ_{FM}.
- μ satisfy the Cheeger's inequality with constant D^μ_{C}.
- μ satisfy the Poincaré inequality: there exists some $D > 0$ such that for all smooth ϕ : $\mathbb{R}^n \to \mathbb{R}$ satisfying $\int \phi \,\mathrm{d}\mu = 0$, we have

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

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

Furthermore, $D^{\mu}_{\text{exp}} \simeq D^{\mu}_{\text{FM}} \simeq D^{\mu}_{\text{P}} \simeq D^{\mu}_{\text{C}}$.

With this theorem and proposition 4.2 in mind, it is clear that the KLS conjecture can be instead formulated with any of these inequalities instead. We also showcase one of these formulations here using the Poincaré inequality:

Conjecture 4 (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.

We remark that isotropic measures satisfy the normalization condition in conjecture 1. Indeed, if $T : \mathbb{R}^n \to \mathbb{R}$ is a 1-Lipschitz linear function, i.e. is of the form $v \mapsto w^T v + d$ for some $w \in S^{n-1}$ and $d \in \mathbb{R}$, then we have

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

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

4.4 Reduction of KLS to thin-shell

We will now present a proof of theorem 6. 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 6 is to control $\text{Var}_{\mu}[\phi]$ by a logarithmic factor of $\text{Var}_{\mu}[\|\cdot\|]$. As translating the barycenter of μ does not affect its variance, we may assume μ has its barycenter $\overline{\mu}$ at the origin. Furthermore, 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. As a result, the region outside said compact set only contributes a bounded amount (in fact, it decreases in n) to the variance, and does not affect our computation (c.f. [Kla06]). Thus, we also have

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

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 $\text{Var}_{\mu}[\phi] = \text{Var}[\phi(a_\infty)]$ where $a_\infty \sim \mu$. Then, for all $t > 0$, by the law of total variance and the martingale property we have

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

where we introduce the notation $\text{Var}[X \mid \mathscr{G}] := \mathbb{E}[(X - \mathbb{E}[X \mid \mathscr{G}])^2 \mid \mathscr{G}]$ 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 $t \geq 0$. 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 the first term, 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 (4) 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 $A_t \to 0$ point-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 (2) 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 $\|A_t\|_{\mathrm{op}} = \max_{i=1,\cdots,n}\lambda_i(t) = \|(\lambda_i(t))_{i=1}^n\|_\infty$ 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 $i = 1, \cdots, n$ and thus we have $\Phi^\alpha(t) = \sum_{i=1}^n \lambda_i(t)^\alpha$. Again, to proceed, we will attempt to compute $\mathrm{d}\Phi^\alpha(t)$ at some $t = t_0 > 0$ utilizing the following simple lemma.

Lemma 4.4. If $A = [a_{ij}]$ is a diagonal matrix with distinct eigenvalues $\lambda_i, \cdots, \lambda_n$, then for all $i, j, k, l, m \in 1, \cdots n$, 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 $e_1, \cdots, e_n$ 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 $\xi_{ij} = \int \langle e_i, (x - a_t)^{\otimes 2} e_j \rangle (x - a_t) \mu_t(\mathrm{d}x)$. Thus, combining this with lemma 4.4, 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 $\mathrm{d}[\lambda_i(t)]_{t_0} = \|\xi_{ii}\|^2 \mathrm{d}t$.
Again applying Itô's formula, we may finally compute

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

where the first inequality holds as

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

Thus, we have shown

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

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 (16)), we may assume without loss of generality that $\|A_t\|_{\mathrm{op}} \geq 1$. Thus, applying the reverse Cauchy-Schwarz inequality to equation (30), we have

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

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

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

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 (16), 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 we have

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

Combining this with equation (24), 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 (the general computation is similar albeit much more technical), let us ignore the stochastic term in equation (31) and regard it as an ODE. Then, by Gronwall's inequality, if we can find some constant K such that $K_t \leq K$ for all $t \leq \tau$, we have the bound

$$S_t \leq n e^{2\alpha^2 K t}.$$

Thus, substituting σ into the above, we have

$$2^{\alpha} = S_{\sigma} \leq n e^{2\alpha^2 K \sigma}$$

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 $\mathbb{E}[\tau^{-1}] \leq 10K \log n$. Of course, this deduction only holds while ignoring the stochastic term $\langle v_t, \mathrm{d}W_t \rangle$. Nonetheless, this is justified as one can show that $\|v_t\|_2$ is bounded $\alpha \Phi^{\alpha}(t)$ and so the same analysis holds by applying the stochastic Gronwall's inequality (c.f. second part of lemma 34 in [LV18]).

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

Lemma 4.5 (Lemma 1.6 in [Eld13])**.** Denoting C_{TS}^n as in theorem 6, there exists a constant C such that for any log-concave, isotropic probability measure μ, we have

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

where $\{e_1, \cdots, e_n\}$ 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

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

We note that we cannot direct apply lemma 4.5 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

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

we have

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

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

At last, combining equation (33) and (32), we have

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

implying there exists a constant $R > 0$ such that for all 1-Lipschitz ϕ, $\sqrt{\text{Var}_\mu[\phi]} \leq RC_{\text{TS}}^n \log n$, i.e. μ is $RC_{\text{TS}}^n \log n$-concentrated and so, $C_{\text{con}}^n \leq RC_{\text{TS}}^n \log n$ 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 log-concave measures. The log-Sobolev inequality is a class of inequalities central to the concentration of measures and has applications to bounding Markov mixing times (c.f.). We will in particular...

TODO

5.1 Entropy and log-Sobolev inequalities

Heuristically, similar to that of the variance, the entropy of a random variable is a measure of its uncertainty or randomness. Formally, the entropy is defined as the following.

Definition 5.1 (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 can be then formulated as the following.

Definition 5.2 (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

Motivation and Herbst.

We will in this section derive a bound for the log-Sobolev constant for all log-concave measures resulting in the following theorem.

Theorem 9 ([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 $\rho_\mu \gtrsim D^{-1}$.

Before presenting Lee and Vempala's proof of above theorem, let us first remark the similarity of the log-Sobolev inequality with that of the Poincaré inequality. As a result, one might first attempt to analyze the log-Sobolev inequality using the same method as prescribed by section 4.4. Given the smooth function $\phi : \mathbb{R}^n \to \mathbb{R}$, let us define the martingale $M_t := \int \phi^2 \mathrm{d}\mu_t$ where (μ_t) is the linear-tilt localization described in section 2.1.1. We observe

$$\begin{aligned}
\mathrm{Ent}_\mu[\phi^2] &= \mathrm{Ent}[M_\infty] = \mathbb{E}[M_\infty \log M_\infty] - \mathbb{E}[M_\infty]\log(\mathbb{E}[M_\infty]) \\
&= \mathbb{E}[\mathbb{E}[M_\infty \log M_\infty \mid \mu_t]] - \mathbb{E}[\mathbb{E}[M_\infty \mid \mu_t]]\log(\mathbb{E}[\mathbb{E}[M_\infty \mid \mu_t]]) \\
&= \mathbb{E}[\mathrm{Ent}[M_\infty \mid \mu_t]] + \mathbb{E}[\mathbb{E}[M_\infty \mid \mu_t]\log(\mathbb{E}[M_\infty \mid \mu_t])] \\
&\qquad - \mathbb{E}[\mathbb{E}[M_\infty \mid \mu_t]]\log(\mathbb{E}[\mathbb{E}[M_\infty \mid \mu_t]]) \\
&= \mathbb{E}[\mathrm{Ent}[M_\infty \mid \mu_t]] + \mathbb{E}[M_t \log M_t] - \mathbb{E}[M_t]\log(\mathbb{E}[M_t]) \\
&= \mathbb{E}[\mathrm{Ent}[M_\infty \mid \mu_t]] + \mathrm{Ent}[M_t]
\end{aligned}$$

Moreover, defining the martingale $N_t := \int \phi^2 \log \phi^2 \mathrm{d}\mu_t$, we observe

$$\begin{aligned}\mathrm{Ent}[M_\infty \mid \mu_t] &= \mathbb{E}[M_\infty \log M_\infty \mid \mu_t] - \mathbb{E}[M_\infty \mid \mu_t] \log(\mathbb{E}[M_\infty \mid \mu_t]) \\ &= \mathbb{E}[N_\infty \mid \mu_t] - M_t \log M_t = N_t - M_t \log M_t \\ &= \mathbb{E}_{\mu_t}[\phi^2 \log \phi^2] - \mathbb{E}_{\mu_t}[]\phi^2] \log \mathbb{E}_{\mu_t}[\phi^2] \\ &= \mathrm{Ent}_{\mu_t}[\phi^2].\end{aligned}$$

Thus, we can express the entropy of ϕ^2 with respect to μ as

$$\mathrm{Ent}_\mu[\phi^2] = \mathrm{Ent}[M_t] + \mathbb{E}[\mathrm{Ent}_{\mu_t}[\phi^2]]. \tag{34}$$

With this expression in mind, we may now attempt to bound each term individually while varying t. Similar to the method in the case of the Poincaré inequality, in which we were able to bound the term $\mathrm{Var}_{\mu_t}[\phi]$ by using the Brascamp-Lieb inequality, we will bound $\mathrm{Ent}_{\mu_t}[\phi^2]$ by using an inequality by Bobkov and Ledoux.

Lemma 5.1 (Bobkov-Ledoux, [BL00]). Let $V : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable strictly convex function such that for all $v \in \mathbb{R}^n$, the map $x \mapsto \langle V''(x)v, v\rangle$ is concave. Then, for all smooth functions $\phi : \mathbb{R}^n \to \mathbb{R}$, we have

$$\mathrm{Ent}_\nu[\phi^2] \le 3\mathbb{E}_\nu[\langle (V'')^{-1} \nabla\phi, \nabla\phi\rangle]$$

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

Hence, by substituting μ_t for ν in the above lemma, where by equation (4) we have that $V'' = \frac{t}{2}\mathrm{id}_n$, the map $x \mapsto \langle V''(x)v, v\rangle$ is concave and hence,

$$\mathrm{Ent}_{\mu_t}[\phi^2] \le 6t^{-1}\mathbb{E}_{\mu_t}[\|\nabla\phi\|^2].$$

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 Jan Brascamp and Elliott H Lieb. On extensions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976.

[BL00] Sergey Bobkov and M. Ledoux. From brunn-minkowski to brascamp-lieb and to logarithmic sobolev inequalities. *Geometric and Functional Analysis*, 10, 12 2000.

[Bl05] Viktor Blåsjö. The isoperimetric problem. *The American Mathematical Monthly*, 112(6):526–566, 2005.

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

[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.

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

[Lev17] David Asher Levin. *Markov chains and mixing times / David A. Levin, Yuval Peres ; with contributions by Elizabeth L. Wilmer.* American Mathematical Society, second edition. edition, 2017.

[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.