Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains

July 30, 2025

A report on the article by Yuansi Chen and Ronen Eldan

Authors: Matthieu Dagès, Vincent Counathe

Contents

1 Introduction

1.1 Why is sampling important?

Sampling from complex probability distributions is a fundamental problem in mathematics, computer science, and statistics. A few applications (and open questions) are given below.

One of the main challenges lies in efficiently sampling from distributions defined on large or high-dimensional spaces. In many cases, direct sampling is computationally infeasible due to the complexity of the distribution or the dimensionality of the space. Therefore, efficient algorithms for sampling are useful.

Applications in Machine Learning and Data Science

- **Generative Models:** Techniques like diffusion models, GANs, and VAEs rely on efficient sampling to generate realistic data.
- **Bayesian Inference:** Sampling algorithms, such as Markov Chain Monte Carlo (MCMC), are used for approximating posterior distributions in Bayesian models.
- **High-Dimensional Optimization:** Sampling is used to understand the geometry of loss landscapes, especially in neural networks.

Intersection with Statistical Physics

- Sampling from high-dimensional distributions is analogous to studying the equilibrium properties of physical systems (e.g., spin glasses).

High-Dimensional Challenges

- In modern problems, distributions often live in extremely high dimensions (e.g., hypercubes).
- Sampling efficiently from such distributions while maintaining accuracy is a major computational and mathematical challenge.

Open Questions

- **Computational Complexity:** Sampling is tied to hard problems in computer science, such as approximating the partition function or sampling in NP-hard combinatorial spaces.
- **Connections to Optimization:** Why does sampling outperform optimization in some cases? This remains a hot topic in stochastic methods.
- **New Methods for High-Dimensional Sampling:** Designing methods that scale with dimension without losing accuracy is an ongoing challenge.

1.2 Markov chains as a tool for sampling

Markov chains provide a powerful and widely used framework for sampling from complex probability distributions. The key idea is to design a Markov chain whose stationary distribution matches the target distribution of interest. By running the chain for a sufficient amount of time, we can obtain samples that approximate the target distribution (more on stationary distributions and sufficient conditions later).

The performance of a Markov chain-based sampling algorithm is closely related to its *mixing time*, which measures how quickly the chain converges to its stationary distribution. Establishing precise bounds on the mixing time is therefore essential to ensure the efficiency and reliability of these algorithms.

1.3 Objectives of this report

This report aims to provide a pedagogical exploration of a recent unifying framework introduced in the paper by Yuansi Chen and Ronen Eldan. The framework bridges two major techniques for analyzing mixing times:

- **Spectral independence**, which has proven effective in discrete settings such as graphical models.
- **Stochastic localization**, a technique that has been applied in both continuous and discrete settings.

The paper is particularly noteworthy because it:

- Unifies these two seemingly unrelated techniques under a single framework.
- Provides simplified proofs of existing results without relying on high-dimensional expanders (HDX), which are technically challenging.
- Extends the applicability of these techniques.

2 Fundamental Concepts

2.1 Markov Chains and Stationary Distributions

A *Markov chain* is a sequence of random variables $X_0, X_1, X_2, \ldots$ where the distribution of each state X_n depends only on the previous state X_{n-1}.

Formally, we define a Markov chain by its *transition matrix (or operator)* P, which provides the probability of moving from one state to another. If we denote the state space of the chain by $\mathcal{S}$ (which can be discrete or continuous), then the transition matrix P is a function that assigns to each pair of states $x, y \in \mathcal{S}$ the probability of moving from x to y, that is, $P(x, y) = \mathbb{P}(X_{n+1} = y \mid X_n = x)$.

One of the key concepts in the study of Markov chains is the *stationary distribution*. A distribution ν over the state space $\mathcal{S}$ is said to be stationary for a Markov chain with transition operator P if it satisfies the following equation:

$$P\nu = \nu$$

In other words, the stationary distribution is invariant under the action of the transition matrix.

Sufficient conditions for the existence of a stationary distribution are:
- Irreducibility: Every state can be reached from every other state in a finite number of steps.
- Aperiodicity: The chain does not get "stuck" in cycles, meaning the greatest common divisor of the lengths of all possible return times to any state is 1.
- Positive recurrence: Every state is visited infinitely often with probability 1, and the expected return time to any state is finite.

When these conditions hold, the Markov chain has a unique stationary distribution, and the chain will converge to this distribution as $n \to \infty$, regardless of the initial state.

2.2 Mixing Times and Their Significance

The *mixing time* of a Markov chain provides a measure of how quickly the chain converges to its stationary distribution from an arbitrary initial distribution.

Given an initial distribution μ, which is absolutely continuous with respect to a stationary distribution ν, consider the total-variation mixing time as follows:

$$t_{\text{mix}}(P, \epsilon; \mu) = \min \left\{ t > 0 \mid \sup_{A \subset \mathcal{S}} \left| P^t[\mu](A) - \nu(A) \right| \leq \epsilon \right\},$$

where $P^t[\mu]$ represents the distribution of the chain at time t when starting from the initial distribution μ, and the total variation distance between the distribution of the chain at time t and the stationary distribution ν is bounded by ϵ for all measurable sets $A \subset \mathcal{S}$. This provides a precise quantitative measure of how long it takes for the chain to become "close" to the stationary distribution.

Moreover, we define the *global mixing time* as the maximum mixing time over all possible initial states. Specifically, for any $x \in \mathcal{S}$, let δ_x denote the Dirac measure at x, i.e., $\delta_x(A) = 1_{\{x \in A\}}$. The global mixing time is then defined as:

$$t_{\text{mix}}(P, \epsilon) = \max_{x \in \mathcal{S}} t_{\text{mix}}(P, \epsilon, \delta_x)$$

In practical terms, this tells us how many steps we need to run the Markov chain to obtain samples that are close to the target stationary distribution.

3 Existing Techniques for Mixing Bounds

3.1 Our setting

The authors consider the following framework: a state-space Ω (for simplicity one can assume $\mathbb{R}^n$ or the hypercube $\{-1,+1\}^n$).

The main objective is to sample for a probability measure ν. For this, we want to construct a reversible Markov chain that will have ν as stationary measure.

Taking the functional equalities point of view, standard Markov chain analysis relates the speed of convergence to the stationary measure ν to the spectral gap $\lambda(P)$, defined as $\lambda(P) := 1 - \max_{\phi, \int \phi d\nu = 0} \frac{\int \phi P \phi d\nu}{\int \phi^2 d\nu}$, with $\phi : \Omega \to \mathbb{R}$.

If we have a lower bound for the spectral gap, we get the following upper bound for the mixing time:

$$t_{\text{mix}}(P, \varepsilon) \leq \lambda_{\text{gap}}(P)^{-1} \left(\log\left(\frac{1}{\eta}\right) + \log\left(\frac{1}{\varepsilon}\right) \right),$$

where $\eta = \min_x \nu(x)$.

This upper bound is only useful when the measure assigns positive mass to all points. Usually η is of order e^{cn}, which means we get a poly-(n) order upper-bound.

Hence, to get polynomial time mixing time, getting a lower bound on the spectral gap can be a sufficient condition to get polynomial time mixing time, whatever the starting point of the Markov Chain.

3.2 Obstacles to fast mixing and sampling via local updates

This section aims to be more on the intuition side, to try to understand obstacles to fast mixing times and how they may relate the covariance structure of the high-dimensional measure.

Let us consider Markov chains that do local updates only (for computational simplicity).

This means, if we consider a graph, then, at each step, we only update one coordinate of the graph, while keeping the other coordinates fixed.

One case is Glauber dynamics. For example, fix ν a probability on the hypercube in dimension n. Given x a configuration in the hypercube, first pick a coordinate $i \in 1...n$ uniformly at random, then generate a point y from the restriction of ν to the hypercube minus the i^{th} coordinate.

In other words,

$$P_{x\to y} = \frac{1}{n}\mathbf{1}_{x\sim y}\frac{\nu(y)}{\nu(x)+\nu(y)}$$

Such markovian dynamics converge to the measure ν (by reversibility of (P,ν)).

One can ask oneself what are the obstacles to fast mixing in that case.

Let us consider Glauber dynamics on the Ising model. In the low temperature regime (high β), adjacent points want to point in the same direction. Thus, in a very low temperature regime, if one state is positive, then this will still have a strong correlation with the state of far-away sites.

If one considers the average spin of a global configuration, there are two very likely configuration types, as β tends to $+\infty$. A large mass will be given two configurations with average spins close to $+1$ and those close to -1, and low mass for average spins close to 0.

This has strong implications for local dynamics like Glauber dynamics. If one wants to sample efficiently, one would want to be able to go from a state with an average spin close to -1 to a state with average spin close to $+1$ (since they are the two most likely configurations), but with a very high β, the Glauber dynamcics will take exponential time to transition from the first state to the second and will get trapped in one of the two macro-states.

This comes as natural way to see how obstacles to fast mixing may arise.

In a way that we will not detail too much in this report, this has to do with the covariance matrix or operator of the measure we wish to sample from. Intuitively, when $\|\mathrm{Cov}(\nu)\|_{\mathrm{op}} \gg 1$, this allows for different configurations that are very hard to transition locally from to another in polynomial time.

In this case, the local property of Glauber dynamics fails to capture the long range-correlations between states, which should be reflected in our sampling strategy and process.

3.3 Localisation schemes

3.3.1 Definitions

The key contribution by the authors is to change the point of view slightly away from Markov chains. To find conditions for fast mixing (i.e in our case, a lower bound for the spectral gap), is to actually consider a stochastic process of measures rather than a Markov chain directly.

Formally the authors define a localization process on Ω as a measure-valued stochastic process $(\nu_t)_{t\geq 0}$ which satisfies the following properties:

(P1) Almost surely, ν_t is a probability measure on Ω for all t.

(P2) For every measurable $A \subset \Omega$, the process $t \mapsto \nu_t(A)$ is a martingale.

(P3) For any measurable $A \subset \Omega$, the process $\nu_t(A)$ almost surely converges to either 0 or 1 as $t \to \infty$.

First, we remark that the localisation process is a process of random measures, starting from the measure which me aim to sample.

Second, this process can be defined both in continuous and discrete times (one can always take a constant process on intervals of the form $[k, k+1)$ for $k \in \mathbb{Z}$.

The authors then introduce a localization scheme on Ω as a mapping that assigns to each probability measure $\nu \in \mathcal{M}(\Omega)$ a localization process $(\nu_t)_{t\geq 0}$ which satisfies $\nu_0 = \nu$.

If L is a localization scheme and $(\nu_t)_{t\geq 0} = L(\nu)$, then we say that $(\nu_t)_{t\geq 0}$ is the localization process associated with ν via the localization scheme L.

We thus get for each measurable $A \subset \Omega$, the localisation scheme starting at ν defines a martingale that is valued in $[0, 1]$, and thus bounded in L^p for all p, and thus converges almost surely. Thus, the stochastic process converges almost surely to a measure ν_∞.

The three properties of the definition ensure that $\nu_\infty = \delta_X$, where $X \sim \nu$. The limiting object is thus a random probability measure, which ensure that after sufficient time, the stochastic process yields a close sample of ν.

We remark that this localisation scheme theoretically provides a sampling scheme. We could run this scheme algorithmically, until it converges to a point. But this may prove to be inefficient, since in most cases it would be computationally costly to store the whole measure, in particular in a high-dimensional setting. The measure would live in a space that is exponentially large in the dimension of Ω. Interestingly, Montanari, El Alaoui and Sellke [AM21] propose an algorithmic stochastic localization framework to sample from the Sherrington-Kirkpatrick Gibbs measure.

What the authors do instead is associate a canonical markov chain to the localisation scheme, for which the stationary probability will be the probability to sample from.

3.3.2 Markov Chain associated to the localization scheme

For a deterministic time, but also for a stopping time τ, one can associate in a very natural way a localisation scheme with a sampling algorithm, via the following Markov chain:

$$P_{x\to A} = \mathbb{E}\left[\frac{\nu_\tau(x)\nu_\tau(A)}{\nu(x)}\right], \quad \forall x \in \Omega,\ A \subset \Omega. \tag{1}$$

We can see that the operator defined in (1) is the transition kernel of a reversible Markov chain whose stationary measure is ν.

Proof. By property (P2), we get $\mathbb{E}[\nu_\tau] = \nu$, which means that for ν-almost every $x \in \Omega$,

$$\mathbb{E}\left[\frac{\nu_\tau(x)\nu_\tau(\Omega)}{\nu(x)}\right] = \mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\right] = 1,$$

so that $P_{x\to\cdot}$ is indeed a probability measure. It is evident from the definition that for all $A, B \subset \Omega$,

$$\int_A P_{x\to B}\, d\nu(x) = \int_A \mathbb{E}\left[\frac{d\nu_\tau(x)}{d\nu(x)}\nu_\tau(B)\right] d\nu(x) = \mathbb{E}[\nu_\tau(A)\nu_\tau(B)] = \int_B P_{y\to A}\, d\nu(y),$$

hence the Markov chain is reversible and has stationary measure ν.

3.3.3 Spectral gap

If a markov chain has transition kernel P, associated to a stochastic localisation scheme $(\nu_t)_t$, then one can show that the spectral gap satisfies:

$$\lambda(P) = \inf_{\phi:\Omega\to\mathbb{R}} \frac{\mathbb{E}\left[\mathrm{Var}_{\nu_\tau}[\phi]\right]}{\mathrm{Var}_\nu[\phi]}. \tag{2}$$

On an intuitive level, the right hand side shows how much of the initial variance the localization scheme keeps at time τ.

This is an important intuition behind localization schemes. Let us think of the test functions as indicators of a subset A of Ω, the localization scheme zooms in over time to one point of Ω. We want to ensure that at time τ, we still see some variance, i.e we get to a place that still has the variance, where there is still a probability to be or not to be in the subset A.

The interesting fact is that a localisation scheme is not necessarily implementable but it gives rise to a naturally implementable process (the markov chain), for which the convergence speed can be controlled using properties depending on the localisation scheme.

3.4 Approximate conservation of variance

One of the main tools used to prove mixing bounds for the dynamics associated with a localization scheme is linked to a property named as approximate conservation of variance.

In this context, a notion of variance decay can be used in order to prove a spectral gap for log-concave measures.

An interesting insight is a connection between the rate of variance decay and the covariance structure, at least for a certain class of measures.The authors define conservation of variance as follows.

A localisation process $(\nu_i)_i$ is said to satisfy $(\kappa_1, \kappa_2, \ldots)$-variance conservation up to time τ, if we have, for all test function $\phi : \Omega \to \mathbb{R}$:

$$\mathbb{E}[\mathrm{Var}_{\nu_i}[\phi] \mid \nu_{i-1}] \geq (1 - \kappa_i)\mathrm{Var}_{\nu_{i-1}}[\phi], \quad \forall 1 \leq i \leq t.$$

The reason for this definition comes from the fact that we can re-write the spectral gap as a telescopic product in the following way.

$$\frac{\mathbb{E}[\mathrm{Var}_{\nu_t}[\phi]]}{\mathrm{Var}_{\nu}[\phi]} = \mathbb{E}\left[\prod_{i=1}^{t} \frac{\mathrm{Var}_{\nu_i}[\phi]}{\mathrm{Var}_{\nu_{i-1}}[\phi]}\right] = \mathbb{E}\left[\prod_{i=1}^{t} \mathbb{E}\left[\frac{\mathrm{Var}_{\nu_i}[\phi]}{\mathrm{Var}_{\nu_{i-1}}[\phi]} \mid \nu_{i-1}\right]\right] \geq \prod_{i=1}^{t}(1 - \kappa_i).$$

In light of this definition and telescopic equality, an easy consequence is the following proposition which relates variance conservation to the spectral gap of the dynamics associated with the localization process.

Proposition 3.1. *If the localization process $(\nu_t)_t$ satisfies $(\kappa_1, \ldots, \kappa_t)$-approximate variance conservation, then the dynamics given in equation (1) with $\tau = t$ has a spectral gap bounded below by $\prod_{i=1}^{t}(1 - \kappa_i)$.*

We now have sufficient conditions which guarantee a lower bound on the spectral gap and therefore an upper bound on the mixing time of the markov chain.

This framework can be used to prove a result on the mixing time of Glauber dynamics first shown by Anari, Liu and Gharan [ALG20]. The stochastic localisation framework allows for a much simpler proof than the one by the above-mentioned authors, who rely on the theory high-dimensional expanders.

The use of simple martingale arguments is arguably one of the strengths of this contribution, and offers promising applications *(see Applications section).*

4 Localization Schemes: A Unified Framework

For a given measure ν, there are several ways to define a stochastic localisation scheme. Choosing the scheme in a smart way may lead to proving mixing time bounds for the associated Markov Chains.

These different schemes provide a generalisation that encompasses many commonly-used and studied markovian dynamics.

4.1 Linear-Tilt Localization Schemes

For the case of coordinate-by-coordinate localization this argument recovers the spectral independence framework of [ALG20], bypassing the need to use highdimensional expanders.

4.1.1 The coordinate-by-coordinate localization example

We start with a simple example presented by the authors. Consider a probability measure ν defined on the hypercube $\{-1, 1\}^n$. The idea of coordinate-by-coordinate localization is to iteratively refine this measure by conditioning on individual coordinates, one step at a time. Here is how the process works:

1. First, randomly choose a permutation $(k_1, k_2, \ldots, k_n)$ of the indices $[n]$ uniformly at random. This ordering determines the sequence in which coordinates will be revealed.

2. Next, sample a random point $X = (X_1, X_2, \ldots, X_n)$ from the measure ν. This represents the "true" state of the system, which we gradually uncover.

3. For a given time $t \geq 0$, define the measure ν_t as the law of X conditioned on the values of $X_{k_1}, X_{k_2}, \ldots, X_{k_i}$, where $i = \min(\lfloor t \rfloor, n)$. In simpler terms, at time $t < n$, we have revealed the first $\lfloor t \rfloor$ coordinates in the order given by $(k_1, k_2, \ldots, k_n)$.

The sequence $(\nu_t)_{t \geq 0}$ forms a martingale. Each ν_t is still a probability measure, but it becomes increasingly "localized" as more coordinates are revealed. By the time $t \geq n$, all coordinates have been revealed, and ν_t almost surely is a single point—a Dirac measure.

The intuition is that this localization process incrementally reduces uncertainty about the random point X. Initially, ν represents the full measure over all possible points in $\{-1, 1\}^n$. As we condition on one coordinate after another, ν_t becomes progressively more localized.

4.1.2 Linear-tilt localizations

We can now introduce the family of linear-tilt localizations, and exhibit the link with the coordinate-by-coordinate localization example.

The central idea behind linear-tilt localizations is the following: to transition from one measure ν_t to the next measure ν_{t+dt}, the density of the measure is adjusted (or "tilted") by multiplying it with a random linear function.

This framework is general and applies whenever the state space Ω can be naturally embedded in a linear space. Our focus will primarily be on two key cases: 1. $\Omega = \mathbb{R}^n$: The continuous setting of the Euclidean space. 2. $\Omega = \{-1, 1\}^n \subset \mathbb{R}^n$: The discrete hypercube setting.

To be consistent with the authors' notations, for ν a given probability measure on the state space Ω, define its center of mass $b(\nu)$ as

$$b(\nu) := \int_\Omega x\, \nu(dx).$$

Intuitively at each small time increment dt, the measure ν_t evolves to ν_{t+dt} by tilting its density proportionally to a random linear function. This process effectively shifts the mass of the measure ν_t, gradually "localizing" it.

We can now revisit the coordinate-by-coordinate localization scheme introduced above and present an equivalent description. This highlights how the scheme fits into the broader family of linear-tilt localizations.

Let ν be a measure on $\Omega = \{-1, 1\}^n$. Define the following random elements:
- $(k_1, \ldots, k_n)$: A uniform random permutation of $[n]$.
- $(U_1, \ldots, U_n)$: An independent sequence of i.i.d. random variables drawn uniformly from $[-1, 1]$.

1. Initialization: Start with $\nu_0 = \nu$.

2. Inductive Update: For $i = 0, 1, \ldots, n-1$, define the next measure ν_{i+1} via:

$$\nu_{i+1}(x) = \nu_i(x)\,(1 + \langle x - b(\nu_i), Z_i\rangle)\,,$$

where $b(\nu_i) = \int_\Omega x\, \nu_i(dx)$ is the center of mass of ν_i, and Z_i is given by:

$$Z_i = e_{k_i} \cdot \begin{cases} \frac{1}{1+b(\nu_i)_{k_i}} & \text{if } b(\nu_i)_{k_i} \geq U_i, \\ -\frac{1}{1-b(\nu_i)_{k_i}} & \text{if } b(\nu_i)_{k_i} \leq U_i. \end{cases}$$

Here, $e_1, \ldots, e_n$ are the standard basis vectors of $\mathbb{R}^n$, and $b(\nu_i)_{k_i}$ is the k_i-th coordinate of $b(\nu_i)$.

3. Extension to Continuous Time: For $t > 0$, define $\nu_t = \nu_{\lfloor t \rfloor \wedge n}$.

Verification as a Localization Process

- Normalization: To compute ν_{i+1}, the measure ν_i is multiplied by a linear function whose expectation is 1. This ensures that $\nu_i(\Omega) = \nu_{i+1}(\Omega)$.

- Zero-mean property: The random vector Z_i satisfies $\mathbb{E}[Z_i] = 0$, verifying the martingale property of a localization process.

- Pinning interpretation: We can view ν_{i+1} as the pinning of ν_i along the k_i-th coordinate. Specifically, if $x_{k_i} = -\text{sign}(U_i - b(\nu_i)_{k_i})$, then $\langle x - b(\nu_i), Z_i\rangle = -1$ (and the linear factor in the update rule vanishes).

To conclude on this alternative definition of the coordinate-by-coordinate localization:

The sequence $(\nu_t)_{t\geq 0}$ can be viewed as a Markov chain on the space of measures. Each transition involves multiplying the measure's density by a random linear function with slopes whose conditional expectation is 0. The slopes are chosen to ensure that ν_{t+1} becomes a restriction of ν_t.

4.2 Stochastic Localization Driven by a Brownian Motion

Let us now describe the stochastic localization process introduced in [Eld13]. This process is defined in the continuous setting and is driven by a Brownian motion.

Let ν be a probability measure on $\mathbb{R}^n$, and let $(B_t)_{t\geq 0}$ be a standard Brownian motion in $\mathbb{R}^n$, adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$. Additionally, let $(C_t)_{t\geq 0}$ be an $\mathcal{F}_t$-measurable process, where C_t is a positive-definite $n \times n$ matrix for all t.

We define a family of probability measures $(\nu_t)_{t\geq 0}$ by changing the density of ν as follows:

$$\frac{d\nu_t}{d\nu}(x) = F_t(x),$$

where the functions $F_t(x)$ satisfy the stochastic differential equation (SDE):

$$F_0(x) = 1, \quad dF_t(x) = F_t(x)\langle x - b(\nu_t), C_t dB_t\rangle, \quad \forall x \in \mathbb{R}^n.$$

Intuitively, given $b(\nu_t) = \int_{\mathbb{R}^n} x\, \nu_t(dx)$ is the expectation of ν_t, $F_t(x)$ adjusts the measure ν based on the deviation of x from the center of mass $b(\nu_t)$.

Using the definition of ν_t, we have:

$$\int_{\mathbb{R}^n} d\nu_t = \int_{\mathbb{R}^n} F_t(x)\, \nu(dx).$$

From the SDE for $F_t(x)$:

$$dF_t(x) = F_t(x)\langle x - b(\nu_t), C_t dB_t\rangle.$$

Integrating over $\mathbb{R}^n$ and using the definition of $b(\nu_t)$:

$$\int_{\mathbb{R}^n} dF_t(x)\nu(dx) = \int_{\mathbb{R}^n} F_t(x)\langle x - b(\nu_t), C_t dB_t\rangle \nu(dx).$$

Since $\int_{\mathbb{R}^n}(x - b(\nu_t))\, \nu_t(dx) = 0$ by the definition of $b(\nu_t)$, we conclude:

$$\int_{\mathbb{R}^n} d\nu_t = 0.$$

Applying Itô's formula to $F_t(x)$, we have:

$$d\log F_t(x) = \langle x - b(\nu_t), C_t dB_t\rangle - \frac{1}{2}\|C_t(x - b(\nu_t))\|^2 dt.$$

Exponentiating, the density function $F_t(x)$ becomes:

$$F_t(x) = \exp\left(\int_0^t \langle x - b(\nu_s), C_s dB_s\rangle - \frac{1}{2}\int_0^t \|C_s(x - b(\nu_s))\|^2 ds\right).$$

This expression ensures that $F_t(x)$ is non-negative. Thus, ν_t satisfies the first property of a localization process (being a probability measure $\forall t$).

Explicit Form of ν_t

We do not prove it here but one can show that for all $t \geq 0$, the measure ν_t takes the explicit form:

$$\frac{d\nu_t}{d\nu}(x) = \exp\left(Z_t - \frac{1}{2}\langle \Sigma_t x, x\rangle + \langle y_t, x\rangle\right),$$

where:
- $\Sigma_t = \int_0^t C_s^2 ds$,
- $y_t = \int_0^t (C_s dB_s + C_s^2 b(\nu_s) ds)$,
- Z_t is a normalizing constant ensuring that ν_t is a probability measure.

This explicit form is useful for proving concentration inequalities.

As regards applications, the stochastic localization process has proven highly useful in studying log-concave measures in the continuous setting, and discrete distributions, including those over $\{-1, 1\}^n$.

It has been employed to derive concentration inequalities in works such as [Che21; Eld13].

5 Approximate conservation of variance for linear-tilt localizations and spectral independence

In this section, we present a powerful tool for proving approximate variance conservation bounds for linear-tilt localization schemes. Note this derivation recovers the main theorem in the spectral independence framework of [ALG20], without using the theory of high-dimensional expanders.

Let $\Omega \subset \mathbb{R}^n$, and let $(\nu_t)_t$ be a localization process on Ω whose evolution is given by the equation:

$$\nu_{t+1}(x) = \nu_t(x)\big(1 + \langle x - b(\nu_t), Z_t\rangle\big), \quad \forall x \in \Omega, \tag{1}$$

where Z_t is a random vector satisfying $\mathbb{E}[Z_t \mid \nu_t] = 0$.

We start by calculating the variance decay of a test function along this process.

Proposition 5.1. *Given a test function $\phi : \Omega \to \mathbb{R}$, it holds that:*

$$\mathbb{E}\big[\operatorname{Var}_{\nu_{t+1}}[\phi] \mid \nu_t\big] - \operatorname{Var}_{\nu_t}[\phi] = -\langle v_t, C_t v_t \rangle, \tag{2}$$

where

$$v_t := \int_\Omega (x - b(\nu_t))\phi(x)\nu_t(\mathrm{d}x), \quad C_t := \operatorname{Cov}(Z_t \mid \nu_t).$$

Proof. Take $\phi : \Omega \to \mathbb{R}$. We have:

$$\begin{aligned}
\mathbb{E}\left[\operatorname{Var}_{\nu_{t+1}}[\varphi] \mid \nu_t\right] &= \mathbb{E}\left[\int_\Omega \varphi(x)^2 \nu_{t+1}(dx) \mid \nu_t\right] - \mathbb{E}\left[\left(\int_\Omega \varphi(x)\nu_{t+1}(dx)\right)^2 \mid \nu_t\right] \\
&= \mathbb{E}_{\nu_t}\left[\varphi^2\right] - \mathbb{E}\left[\left(\int_\Omega \left(1 + \langle x - \mathbf{b}\left(\nu_t\right), Z_t\rangle\right)\varphi(x)\nu_t(dx)\right)^2 \mid \nu_t\right] \\
&= \mathbb{E}_{\nu_t}\left[\varphi^2\right] - \mathbb{E}_{\nu_t}[\varphi]^2 - \mathbb{E}\left[\left(\int_\Omega \langle x - \mathbf{b}\left(\nu_t\right), Z_t\rangle\, \varphi(x)\nu_t(dx)\right)^2 \mid \nu_t\right], \\
&= \operatorname{Var}_{\nu_t}[\varphi] - \operatorname{Var}\left[\left\langle \int_\Omega \left(x - \mathbf{b}\left(\nu_t\right)\right)\varphi(x)\nu_t(dx), Z_t \right\rangle \mid \nu_t\right]
\end{aligned}$$

where we first use the fact that ν_t is a martingale and then that $\mathbb{E}[Z_t \mid \nu_t] = 0$. Using the definitions of v_t and C_t, the right-hand side simplifies to:

$$\operatorname{Var}_{\nu_t}[\phi] - \langle v_t, C_t v_t \rangle.$$

□

To establish an approximate variance conservation bound, we need an upper bound on the term $\langle v_t, C_t v_t \rangle$. By Cauchy-Schwarz, we have:

$$\begin{aligned}
\langle v_t, C_t v_t \rangle &= \left\| \int_\Omega C_t^{1/2}(x - b(\nu_t))\phi(x)\nu_t(\mathrm{d}x) \right\|^2 \\
&\leq \sup_{\|\theta\|=1} \left(\int_\Omega \langle C_t^{1/2}(x - b(\nu_t)), \theta\rangle^2 \nu_t(\mathrm{d}x) \right) \operatorname{Var}_{\nu_t}[\phi] \\
&= \|C_t^{1/2} \operatorname{Cov}(\nu_t) C_t^{1/2}\|_{\mathrm{op}} \operatorname{Var}_{\nu_t}[\phi].
\end{aligned}$$

Combining this with Equation (2), we have:

$$\mathbb{E}\big[\operatorname{Var}_{\nu_{t+1}}[\phi] \mid \nu_t\big] \geq \operatorname{Var}_{\nu_t}[\phi] \left(1 - \|C_t^{1/2} \operatorname{Cov}(\nu_t) C_t^{1/2}\|_{\mathrm{op}}\right). \tag{3}$$

This is where the covariance structure plays a role in variance conservation.

5.1 Application to Coordinate-by-Coordinate Localization

Consider the coordinate-by-coordinate localization scheme. Using the definition of $b(\nu_t)$, it holds that:

$$\begin{aligned}
(n - t) \operatorname{Cov}\left[Z_t \mid \nu_t\right]_{i,i} &= \frac{1}{\left(1 + \mathbf{b}\left(\nu_t\right)_i\right)^2} \frac{1 + \mathbf{b}\left(\nu_t\right)_i}{2} + \frac{1}{\left(1 - \mathbf{b}\left(\nu_t\right)_i\right)^2} \frac{1 - \mathbf{b}\left(\nu_t\right)_i}{2} \\
&= \frac{1}{1 - \mathbf{b}\left(\nu_t\right)_i^2} = \left(\operatorname{Cov}\left(\nu_t\right)_{i,i}\right)^{-1},
\end{aligned}$$

using the fact that

$$\int_\Omega x_i^2 \, \nu_t(dx) = 1.$$

If we denote D_t the diagonal matrix with diagonal entries match those of $\mathrm{Cov}(\nu_t)$, we find:

$$\|C_t^{1/2} \mathrm{Cov}(\nu_t) C_t^{1/2}\|_{\mathrm{op}} = \frac{1}{n-t} \|D_t^{-1/2} \mathrm{Cov}(\nu_t) D_t^{-1/2}\|_{\mathrm{op}}.$$

Define the correlation matrix:

$$\mathrm{Cor}(\nu) := \mathrm{diag}(\mathrm{Cov}(\nu))^{-1/2} \mathrm{Cov}(\nu) \, \mathrm{diag}(\mathrm{Cov}(\nu))^{-1/2},$$

where $\mathrm{diag}(\cdot)$ is the diagonal matrix obtained by setting all the off-diagonal entries to 0.

Consider also the influence matrix $\Psi(\nu)$, defined as:

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

$\Psi(\nu)$ and $\mathrm{Cor}(\nu)$ are related by the following inequality:

$$\|\mathrm{Cor}(\nu)\|_{\mathrm{op}} = \rho(\Psi(\nu)) \leq \|\Psi(\nu)\|_{\mathrm{op}}.$$

Which, using (3), allows to finally conclude:

$$\mathbb{E}\big[\mathrm{Var}_{\nu_{t+1}}[\phi] \mid \nu_t\big] \geq \mathrm{Var}_{\nu_t}[\phi] \left(1 - \frac{\rho(\Psi(\nu_t))}{n-t}\right).$$

Let $u \in \{-1, 0, 1\}^n$. The u-pinning of a measure ν, denoted $R_u\nu$, is defined as the restriction of ν to the sub-cube S_u, where

$$S_u := \{x \in \{-1, 1\}^n \mid x_i u_i \geq 0, \forall i \in [n]\}.$$

Note that under coordinate-by-coordinate localization, ν_t can be written as $R_u\nu$ for some $u = u(t)$. The condition that $\|\Psi(R_u\nu_t)\|_{\mathrm{OP}}$ is uniformly bounded in u is referred to as spectral independence.

Substituting this into Equation (3) yields the main result of the spectral independence framework in [ALG20]:

Theorem 5.1. *(A reformulation of [ALG20], Theorem 1.3) Let ν be a measure on $\{-1, 1\}^n$ such that for all $u \in \{-1, 0, 1\}^n$,*

$$\rho(\Psi(R_u\nu)) \leq \eta |u|_1.$$

Then the spectral gap of the k-Glauber dynamics on ν is at least

$$\prod_{i=0}^{n-k-1} \left(1 - \frac{\eta_i}{n-i}\right).$$

Remark. On the differences between the above theorem and [ALG20], Theorem 1.3: in [ALG20], $\tilde{\eta}_i$ is taken to be the operator norm of the matrix $\Psi(R_i\nu) - I_n$, rather than $\Psi(\nu)$. Specifically, their result shows that the spectral gap is bounded from below by the expression

$$\frac{1}{n}\prod_{i=0}^{n-k-1}\left(1-\frac{\tilde{\eta}_i}{n-i-1}\right),$$

where $\tilde{\eta}_i = \eta_i - 1$. This extra factor of $\frac{1}{n}$ accounts for the replacement of η_i with $\tilde{\eta}_i = \eta_i - 1$.

In their proof, Eldan and Chen bypassed the need for concepts like high-dimensional expanders or the up-down walk. The inequalities involving different levels of the up-down walk were replaced with an application of the Cauchy-Schwarz inequality.

6 Applications of localization schemes: Glauber dynamics for Ising models

We begin by presenting a general theorem that provides a sufficient condition for MLSI (Modified Log-Sobolev Inequality). This corresponds in a sense to a generalisation of the spectral gap property, using entropy rather than variances, and can lead to optimal bounds for mixing times. The idea behind the proofs are similar to the case of the spectral gap.

In the case of a Markov chain associated to a localisation scheme, the MLSI coefficient $\rho_{LS}(P)$ can be shown to satisfy:

$$\rho_{LS}(P) \geq \inf_{\phi:\Omega\to\mathbb{R}^+} \frac{\mathbb{E}\left[\mathrm{Ent}_{\nu_\tau}[\phi]\right]}{\mathrm{Ent}_\nu[\phi]}. \tag{2}$$

Again, providing a lower bound for $\rho_{LS}(P)$ yields an upper bound for the mixing time of the associated Markov chain.

This theorem also recovers two important results as special cases. The first result, shown in [EKZ21; Ana+21], establishes the mixing time for Ising models whose interaction matrix has a bounded operator norm. This result is particularly relevant for the Sherrington-Kirkpatrick model at high temperatures. The second result applies to Ising models in the uniqueness regime and improves upon the main theorem of [Che+21]. Both of these results rely on the following theorem.

Theorem 6.1. *Let ν be a probability measure on $\{-1,1\}^n$, and let J be a positive-definite $n \times n$ matrix. For any $0 \leq \lambda \leq 1$ and any $v \in \mathbb{R}^n$, define the probability measure $\mu_{\lambda,v}$ by*

$$\frac{d\mu_{\lambda,v}}{d\nu}(x) \propto \exp\left(-\lambda\langle x, Jx\rangle + \langle v, x\rangle\right),$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product. We assume the following conditions:
1. For some function $\alpha : [0,1] \to \mathbb{R}^+$,

$$\|Cov(\mu_{\lambda,v})\|_{OP} \leq \alpha(\lambda), \quad \forall \lambda \in [0,1], v \in \mathbb{R}^n.$$

2. For some constant $\epsilon > 0$,

$$\rho_{LS}(P_{GD}(\mu_{1,v} z)) \geq \epsilon, \quad \forall v \in \mathbb{R}^n,$$

where $P_{GD}(\cdot)$ denotes the transition kernel of the Glauber dynamics. Then, the following inequality holds for the measure ν:

$$\rho_{LS}(P_{GD}(\nu)) \geq \epsilon \exp\left(-2\|J\|_{OP} \int_0^1 \alpha(\lambda)\, d\lambda\right).$$

This theorem is especially useful for measures with quadratic potentials, such as Ising models. For an $n \times n$ matrix J and a vector $v \in \mathbb{R}^n$, consider the Ising measure $\nu_{J,v}$ defined by

$$\nu_{J,v}(\{x\}) \propto \exp\left(\langle x, Jx \rangle + \langle x, v \rangle\right),$$

where v denotes the external field. If we apply the theorem above to the measure $\nu = \nu_{J,v}$, we observe that the measure $\mu_{1,v}$ defined above is simply a product measure. This means that the condition on ρ_{LS} is satisfied with $\epsilon = \frac{1}{n}$. Therefore, the remaining task is to verify the condition on $\|\mathrm{Cov}(\mu_{\lambda,v})\|_{\mathrm{OP}}$, which ensures that the operator norm of the covariance is bounded in the way specified.

Proof. To prove this theorem, we consider the localization process $(\nu_t)_t$ corresponding to the measure ν, which is obtained via stochastic localization. In particular, we choose $C_t = (2J)^{1/2}$ up to time $t = 1$. We then consider the transition kernel $P_{\mathrm{GD}}(\nu_1)$, which is the random transition kernel associated with the measure ν_1 via the Glauber dynamics. The goal is to apply Theorem 47 in the original paper to obtain a lower bound on $\rho_{\mathrm{LS}}(P_{\mathrm{GD}}(\nu))$.

ν_t has the form

$$\nu_t(x) \propto \exp\left(-t\langle x, Jx \rangle + \langle y_t, x \rangle\right) \nu(x),$$

for some stochastic process y_t. This shows that ν_t is of the form μ_{t,y_t}, and in particular, $\nu_1 = \mu_{1,y_1}$. By hypothesis $\rho_{\mathrm{LS}}(P_{\mathrm{GD}}(\nu_1)) \geq \epsilon$, satisfying the second condition of Theorem 47.

Also by hypothesis, we know that the operator norm of the covariance satisfies

$$\|\mathrm{Cov}(\nu_t)\|_{\mathrm{OP}} \leq \alpha(t), \quad \forall t \in [0,1].$$

Using Lemma 40 in the original paper with the choice $A = \alpha(t) I_n$ and $C = (2J)^{1/2}$, we conclude that ν_t is α-entropically stable with respect to the potential function $\psi(x,y) = \frac{1}{2}\|(2J)^{1/2}(x-y)\|^2$, with $\alpha = 2\|J\|_{\mathrm{OP}}\alpha(t)$.

Now, applying Proposition 39 in the original paper, we obtain the entropy conservation inequality

$$\mathbb{E}[\mathrm{Ent}_{\nu_1}[f]] \geq e^{-2\|J\|_{\mathrm{OP}} \int_0^1 \alpha(t)\, dt}\, \mathrm{Ent}_{\nu}[f].$$

This shows that condition 1 of Theorem 47 is satisfied with $\alpha = e^{-2\|J\|_{\mathrm{OP}} \int_0^1 \alpha(t)\, dt}$. Thus, we conclude that

$$\rho_{\mathrm{LS}}(P) \geq \epsilon \exp\left(-2\|J\|_{\mathrm{OP}} \int_0^1 \alpha(t)\, dt\right).$$

This completes the proof. □

7 Conclusion

To derive mixing time bounds for Markov chains, which are crucial to make sure that sampling from a give measure is possible in polynomial time, the authors have introduced a powerful framework which constructs a Markov chain with the right stationary measure to sample from. By considering this probabilistic framework, rather than the Markov chain from which they are derived, one can provide deep and powerful results.

The authors build on the second author's framework of localization schemes [Eld13] to derive cold start mixing time guarantees for Markov chains in a simple probabilistic framework that completely bypasses the need for high-dimensional expanders.

Stochastic localisation is a powerful framework that has enabled the research community not only to derive simpler proofs for existing mixing time results, but also provide novel results.

The framework is sufficiently broad to encompass a broad range of local dynamics and prove mixing time bounds. Applications are numerous and include Glauber dynamics (e.g. for the Ising model) or sampling from strongly-log concave measures.

It is still an active branch of research. While the localization schemes are broad and theoretically elegant, their practical implementation for specific Markov chains (especially in high-dimensional settings) remains less clear. Concrete algorithms derived from these methods may involve significant computational overhead. This could be an interesting future line of research.

The paper provides theoretical results but lacks an empirical evaluation of how these techniques perform on real-world problems. This is particularly relevant for applications like sampling from Ising models or log-concave measures, where computational comparisons with other methods could be insightful.

References

[ALG20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral independence in high-dimensional expanders and applications to the hard-core model". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*. Los Alamitos, CA: IEEE Computer Society, 2020, pp. 1319–1330.

[AM21] Ahmed El Alaoui and Andrea Montanari. "An information-theoretic view of stochastic localization". In: *arXiv preprint* arXiv:2109.00709 (2021).

[Ana+21] Nima Anari et al. "Entropic independence in high-dimensional expanders: modified log-Sobolev inequalities for fractionally log-concave polynomials and the Ising model". In: *arXiv preprint* arXiv:2106.04105 (2021).

[Che+21] Xiaoyu Chen et al. "Optimal Mixing Time for the Ising Model in the Uniqueness Regime". In: *arXiv preprint* (2021). arXiv: `2111.03034 [math.PR]`.

[Che21] Yuansi Chen. "An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture". In: *Geometric and Functional Analysis* 31.1 (2021), pp. 34–61.

[EKZ21] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. "A spectral condition for spectral gap: fast mixing in high-temperature Ising models". In: *Probability Theory and Related Fields* (2021), pp. 1–17.

[Eld13] Ronen Eldan. "Thin shell implies spectral gap up to polylog via a stochastic localization scheme". In: *Geometric and Functional Analysis* 23.2 (2013), pp. 532–569.