Lecture 21: Localization Schemes

November 28, 2023

In this lecture, we study a significantly more general and powerful framework for proving rapid mixing of Markov chains called *localization schemes* [CE22]. It includes the spectral/entropic independence framework as a special case, but also encompasses the famous *stochastic localization* method of Eldan [Eld13]. The latter has been successfully applied to a myriad of longstanding open problems, with the most notable among them being perhaps the (near-)resolution [Eld13; LV17; LV18; Che21] of the *Kannan-Lovász-Simonovits (KLS) Conjecture* [KLS95] from asymptotic convex geometry. Stochastic localization has also been used recently as a sampling algorithm [AMS22; MW23].[1] Finally, there are additional connections with the statistical physics *renormalization group* method [BBD23]. It is certainly not possible to do justice to stochastic localization in a single lecture. The goal here is to present an overarching framework encompassing all of these as special cases, and focus on applications to Markov chain mixing times.

1 Localization Processes

Definition 1 (Localization Process/Scheme; [CE22])**.** *Let μ be a probability measure on a state space Ω (e.g. $\{\pm 1\}^n, \mathbb{R}^n$). A* localization process for μ *is a sequence of* random *probability measures $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ satisfying the following properties:*

1. ***Initialization:*** *$\mu^{(0)} = \mu$ with probability* 1*.*
2. ***(Iterative) Refinement:*** *For every event $A \subseteq \Omega$, the stochastic process $\{\mu^{(t)}(A)\}_{t\in\mathbb{R}_{\geq 0}}$ on $\mathbb{R}_{\geq 0}$ is a* martingale*. This means that for all $0 \leq T_1 \leq T_2$,*

$$\mathbb{E}\left[\mu^{(T_2)}(A) \mid \left\{\mu^{(t)}(A) : 0 \leq t \leq T_1\right\}\right] = \mu^{(T_1)}(A).$$

   Here, the expectation is w.r.t. the randomness of choosing a sequence of measures $\{\mu^{(t)}\}_{t\in\mathbb{R}_{\geq 0}}$.
3. ***Localization:*** *For all events $A \subseteq \Omega$, $\mu^{(t)}(A)$ converges to either* 0 *or* 1 *almost surely as $t \to \infty$.*

A localization scheme $\mathcal{L}$ *on a state space Ω is a mapping from every probability measure μ over Ω to a localization process for μ.*[2] *We sometimes write $\mathcal{L}(\mu, t)$ for the random measure $\mu^{(t)}$ obtained by running the localization specified by $\mathcal{L}$ on μ up to time t.*

Informally, a localization process is a structured sequence of decompositions of the measure μ. This sequence of decompositions is "iterative" in the sense that for each $0 \leq s \leq t \leq \infty$, the decomposition at "time" t is obtained by applying further decomposition to the random measure $\mu^{(s)}$. The sequential nature underlies the power of this framework.

To make things (perhaps painfully) explicit and to avoid possible confusion regarding the meaning of "$\mathbb{E}$", "randomness of the measure $\mu^{(t)}$", etc., here is an alternative description of a localization process which we will employ occasionally below. For each $t \in \mathbb{R}_{\geq 0}$, there is some index set $\mathcal{I}^{(t)}$, a probability measure $\xi^{(t)}$ over $\mathcal{I}^{(t)}$ called the *mixture measure*, and a collection of probability measures $\left\{\nu_\iota : \iota \in \mathcal{I}^{(t)}\right\}$ over Ω called *component measures*, such that

- $\mu^{(t)}$ is a random probability measure equalling the component ν_ι with probability $\xi^{(t)}(\iota)$, and

[1] However, notably these sampling results only yield $o(n)$ error in Wasserstein distance w.r.t. the Euclidean metric. In particular, they do not yield FPAS for sampling, nor FPRAS for counting.

[2] The notation $\mathcal{L}$ here is not to be confused with the logarithmic Laplace transform.

- $\mu(x) = \mathbb{E}_{\iota \sim \xi^{(t)}}[\nu_\iota(x)]$ for all $x \in \Omega$.[3]

The first initialization criterion mandates that $\mathcal{I}^{(0)}$ is a singleton with $\nu_\iota = \mu$ where ι is the unique element of $\mathcal{I}^{(0)}$. The third localization criterion enforces that at time $t = \infty$, $\mu^{(\infty)}$ is a Dirac measure on Ω. Note that when combined with the initialization and martingale conditions, this means that $\mu^{(\infty)} = \delta_x$ with probability $\mu(x)$. This alternative viewpoint is not quite correct because we have not enforced the martingale property at all, which should intuitively mean that for every $0 \le s \le t \le \infty$, the decomposition at time t "refines" the one at time s; in other words, $\mu^{(t)}$ is obtained by further decomposing $\mu^{(s)}$. Nonetheless, this perspective will sometimes be useful when explaining things at an informal level.

Although we have formulated localization processes in continuous-time, the discrete-time version is already captured by letting $\mu^{(t)}$ be constant on intervals $[k, k+1)$ for all $k \in \mathbb{N}$. Definition 1 is best understood by seeing multiple examples, and how a single localization process can be defined in many (not obviously equivalent) ways. One such localization process, the *coordinate-by-coordinate* localization, is one which we are already very familiar with. Additional examples of localization processes (beyond those mentioned in this lecture) are provided in [CE22].

1.1 The "Formulaic" Perspective

First, we can build localization processes via direct methods. This way of viewing things does not capture the martingale property for a localization process, but is useful for getting an explicit handle on what each $\mu^{(t)}$ is marginally.

Example 1 (Coordinate-by-Coordinate Localization). Let μ be a probability measure $\{\pm 1\}^n$ (or, more generally, $[q]^n$). Let $\mu^{(0)} = \mu$, and let $\mu^{(1)}$ be the random probability measure given by $\mu^{i \leftarrow s}$, where $i \sim [n]$ and $s \sim \mu_i$. Inductively applying this one-step decomposition to $\mu^{(t)}$ to obtain $\mu^{(t+1)}$ for each $t \in \{0, \dots, n\}$, we obtain a discrete-time localization process where

$$\mu^{(t)} = \mathbb{E}_{S \sim \binom{[n]}{t}}\left[\mathbb{E}_{\tau \sim \mu_S}[\mu^\tau]\right], \qquad \forall t \in \{0, \dots, n\}.$$

More explicitly, for each $t \in \{0, \dots, n\}$, the index set $\mathcal{I}^{(t)}$ is given by pairs (S, τ) where $S \in \binom{[n]}{t}$ and τ is a partial assignment supported on S. We have $\xi^{(t)}(S, \tau) = \frac{1}{\binom{n}{t}} \cdot \mu_S(\tau)$, i.e. $\xi^{(t)}$ is a mixture of the marginal distributions μ_S over a uniformly random $S \sim \binom{[n]}{t}$. The component measures are the conditionals of μ based on pinning all coordinates in S. Note that $\xi^{(t)}$ is essentially the measure "μ_t" we defined previously in the context of spectral and entropic independence.

Example 2 (Gaussian Channel Localization). Let μ be a probability measure on $\Omega \subseteq \mathbb{R}^n$. Fix a positive definite "driving matrix" $A \succ 0$. For each $t \in \mathbb{R}_{\ge 0}$, define

$$\mu^{(t)}(x) \propto \mu(x) \cdot \exp\left(-\frac{1}{2}(z_t - tx)^\top \frac{A}{t}(z_t - tx)\right), \qquad \forall x \in \Omega,$$

where z_t "parametrizes" the component distribution $\mu^{(t)}$, and is drawn from the *convolution* of μ with the (independent) Gaussian $\mathcal{N}\left(0, tA^{-1}\right)$:

$$\xi^{(t)}(z) = \mathbb{E}_{x \sim \mu}\left[\mathbb{E}_{y \sim \mathcal{N}(0, tA^{-1})}\left[\mathbf{1}_{z = tx + y}\right]\right].$$

Note that in the definition of $\mu^{(t)}$, one can expand the Gaussian part and note that the term $-\frac{1}{2} z_t^\top \frac{A}{t} z_t$ cancels after normalization. Hence, we can also express $\mu^{(t)}$ as

$$\mu^{(t)}(x) \propto \mu(x) \cdot \exp\left(-\frac{1}{2} t x^\top A x + x^\top A z_t\right), \qquad \forall x \in \Omega. \tag{1}$$

So, we are introducing a Gaussian component in μ which concentrates more and more as we increase t; note that this guarantees that $\mu^{(\infty)}$ is a Dirac mass. We also have an exponential tilt component in the direction Az_t. This is a special case of the *stochastic localization* process introduced in [Eld13], where it is possible to let the choice of "driving matrix" A depend on t.

[3] Note that each ν_ι must be absolutely continuous w.r.t. μ (almost surely).

Example 3 (Negative Fields Localization). Let μ be a probability measure $\{\pm 1\}^n$. For each $t \in \mathbb{R}_{\geq 0}$, define

$$\mu^{(t)}(\sigma) \propto \mu(\sigma) \cdot \exp\left(-t \cdot \left|\sigma^{-1}(+1) \setminus S_t\right|\right), \qquad \forall \sigma \in \{\pm 1\}^n \text{ s.t. } \sigma|_{S_t} \equiv \mathbf{1},$$

where S_t "parametrizes" the component distribution $\mu^{(t)}$, and is drawn from the following mixture distribution:

$$\xi^{(t)}(S) = \sum_{\substack{\sigma \in \{\pm 1\}^n \\ \sigma^{-1}(+1) \supseteq S}} \mu(\sigma) \cdot \left(\frac{1}{1+e^{-t}}\right)^{|S|} \left(\frac{e^{-t}}{1+e^{-t}}\right)^{\left|\sigma^{-1}(+1) \setminus S\right|}, \qquad \forall S \subseteq [n]. \tag{2}$$

In this case, somewhat similar to the coordinate-by-coordinate localization process, the component distributions are obtained from μ by applying exponential tilts. More specifically, we apply an external field which biases the distribution towards -1. To compensate, we pin a special subset of coordinate S_t all to $+1$. We will sometimes write $(\mathcal{T}_{-t\mathbf{1}}\mu)^{S_t \leftarrow +\mathbf{1}}$ for this distribution to make things explicit. Since we're looking at the Boolean cube, in the case $t \to \infty$, the distribution $\xi^{(\infty)}$ concentrates on $\sigma^{-1}(+1)$, which uniquely determines σ and ensures that $\mu^{(\infty)}$ is a Dirac mass. Hence, the process indeed localizes.

1.2 The Statistical Inference Perspective

Most localization processes can be alternatively defined via statistical inference problems. The rough skeleton of this information-theoretic approach is as follows. Let $\mathcal{D}^{(t)}$ denote some class of noisy "observation channels" such that $t = 0$ corresponds to "infinite noise", and $t = \infty$ corresponds to "zero noise". More formally, each $\mathcal{D}^{(t)}$ is some Markov kernel mapping distributions on Ω to distributions on some other auxiliary space $\mathcal{I}^{(t)}$. Our target measure μ is some idealized distribution which we do not have direct access to. However, we have "approximate" access in the following sense:

- "Nature" draws $X \sim \mu$, often called the "signal".
- This sample X is then corrupted by sending it through the channel $\mathcal{D}^{(t)}$, yielding a noisy copy Y_t of X often called the "observation". t is often thought of as a *signal-to-noise ratio (SNR)* parameter.

The inference problem is to recover X from knowledge of Y_t. For the purposes of building a localization process, the observation Y_t induces a conditional measure which we take to be $\mu^{(t)}$. The mixture measure $\xi^{(t)}$ arises from channel $\mathcal{D}^{(t)}$. This perspective is very useful at a conceptual level as it suggests many connections with other areas within statistics and information theory. To translate this formulation into the formulaic approach of Section 1.1, the key is to use Bayes' Theorem. However, we emphasize that this discussion is much too informal; see Remark 1.

Example 4 (Coordinate-by-Coordinate (Cont.)). Continuing Example 1, let μ be a probability measure $\{\pm 1\}^n$ (or, more generally, $[q]^n$). To highlight the "paths" from $\mu^{(0)} = \mu$ to the random Dirac measure $\mu^{(\infty)}$, let $\sigma \sim \mu$ and $(i_1, \ldots, i_n) \sim \mathsf{Unif}\{\text{permutations } \pi : [n] \to [n]\}$. Then take

$$\mu^{(t)} = \mathsf{Law}\left(\sigma \mid \sigma(i_1), \ldots, \sigma(i_t)\right), \qquad \forall t \in \{0, \ldots, n\}.$$

In other words, take the noisy channel $\mathcal{D}^{(t)}$ to be a "homogenized version" of the *erasure channel*: We only reveal a uniformly random t-subset $S \in \binom{[n]}{t}$ of the coordinates of a sample $\sigma \sim \mu$; all other coordinates are "erased". In this set up, the noisy channel $\mathcal{D}^{(t)}$ is precisely the down operator $\mathcal{D}^{n \searrow t}$ we previously defined.

Example 5 (Gaussian Channel (Cont.)). Continuing Example 2, let μ be a probability measure on $\Omega \subseteq \mathbb{R}^n$. Fix a positive definite matrix $A \succ 0$. Let $X \sim \mu$ and let $(B_t)_{t=0}^{\infty}$ be a Brownian motion on $\mathbb{R}^n$ which is independent of X.[4] We define the (continuous-time) *Gaussian channel localization process* by

$$\mu^{(t)} = \mathsf{Law}\left(X \mid tX + A^{-1/2}B_t\right), \qquad \forall t \in \mathbb{R}_{\geq 0}.$$

[4]Recall that, informally, this is a coupled sequence of Gaussians such that $B_t - B_s \sim \mathcal{N}(0, (t-s)\mathsf{Id})$ for all $0 \leq s < t \leq \infty$.

The channel $\mathcal{P}_t$ is just corruption with additive Gaussian noise. This perspective on stochastic localization was first elucidated in [AM22].

Example 6 (Negative Fields (Cont.)). Continuing Example 3, let μ be a probability measure on $\{\pm 1\}^n$. Let $\sigma \sim \mu$, and independently build a random (coupled) sequence $(S_t)_{t\in\mathbb{R}_{\geq 0}}$ of subsets of $[n]$ as follows:

- We initialize with $R_0 = \emptyset$. Throughout, we ensure the sequence is *increasing* in the sense that $S_{T_1} \subseteq S_{T_2}$ for all $0 \leq T_1 \leq T_2 \leq \infty$.
- If the subsets have been sampled up to some time t_ℓ (for some $\ell \in \{0, \dots, n\}$), then for each $i \in [n] \setminus S_{t_\ell}$, let T_i be an independent random variable uniquely defined by

$$\Pr[T_i > y] \stackrel{\mathsf{def}}{=} \exp\left(-\int_{t_\ell}^{y} \Pr[\sigma(i) = +1]\, ds\right),$$

  where Pr is w.r.t. the measure $(\mathcal{T}_{-s\mathbf{1}}\mu)^{S_{t_\ell} \leftarrow +\mathbf{1}}$ on $\{\pm 1\}^n$ obtained by pinning all coordinates of S_{t_ℓ} to $+1$ and applying an external field of e^{-s} to bias μ further towards -1. Letting

$$t_{\ell+1} \stackrel{\mathsf{def}}{=} \min_{i\in[n]\setminus S_{t_\ell}} T_i \qquad \text{and} \qquad i^* \stackrel{\mathsf{def}}{=} \arg\min_{i\in[n]\setminus S_{t_\ell}} T_i,$$

  we extend the sequence of subsets by letting $S_t = S_{t_\ell} + i^*$ for all $t_\ell < t \leq t_{\ell+1}$. Once we reach t_n, $S_t = [n]$ for all $t > t_n$.

The point is that each S_t is marginally distributed as $\xi^{(t)}$ from Eq. (2), but that the entire sequence is coupled so as to guarantee it is increasing. The first point can be checked directly, and furthermore

$$\mu^{(t)} = \mathsf{Law}(\sigma \mid \sigma_{S_t}).$$

In other words, the noise channel consumes a random $\sigma \sim \mu$, and outputs a random subset of $\sigma^{-1}(+1)$ where each coordinate is added independently with probability $\frac{1}{1+e^{-t}}$. Note that while the law of this random sequence of sets depends on μ, the sequence is generated independently from $\sigma \sim \mu$.

Remark 1. The discussion at the beginning of this subsection is problematic since it doesn't prescribe *how* the noise channels $\mathcal{D}^{(t)}$ are related to one another. For instance, it doesn't capture the crucial property that the sequence of noisy observations $\{Y_t\}_{t\in\mathbb{R}_{\geq 0}}$ should be coupled in a way such that for every $0 \leq s \leq t \leq \infty$, Y_s is obtained from Y_t by adding more noise. Notably, this property is clearly maintained in Examples 4 to 6. The correct way to formalize this measure-theoretically is through *filtrations* and the notion of a *Doob localization scheme*; see Appendix A for further discussion.

1.3 The Stochastic Differential Equation (SDE) Perspective

A third way to build a localization process is to do it "dynamically". We prescribe the "infinitesimal change" between $\mu^{(t)}$ and $\mu^{(t+\,dt)}$ for an infinitesimally small[5] dt (i.e. the "dynamics"), and let this implicitly induce a sequence of random probability measures as the solution of a system of stochastic differential equations. This perspective is useful for calculations since one can take derivatives and bring in the power of *Itô calculus*.

In many useful examples, this infinitesimal change is described by multiplication by an affine function, suitably normalized so that the resulting object is still a probability measure,

$$\frac{\mu^{(t+\,dt)}(x)}{\mu^{(t)}(x)} = 1 + \left\langle x - \boldsymbol{m}\left(\mu^{(t)}\right), dz_t \right\rangle, \tag{3}$$

where $\boldsymbol{m}(\mu)$ denotes the mean of a distribution on a subset of $\mathbb{R}^n$, and z_t is some (possibly random) sequence of vectors satisfying $\mathbb{E}\left[dz_t \mid \mu^{(t)}\right] = 0$.[6] The first term $+1$ and the subtraction

[5] For all purposes in the lecture, it is fine to think of dt as a tiny but strictly positive real number.

[6] The second manipulation only really makes sense if the localization process is genuinely in continuous-time and there is some smoothness assumption.

$-\boldsymbol{m}(\mu^{(t)})$ together ensure that $\mu^{(t+\,dt)}$ still integrates to 1. This is technically an infinite system of stochastic differential equations, but under various regularity conditions (which are often satisfied in practice), they nonetheless admit a solution [Eld13; EKZ22]. This special type of localization process is called a *linear tilt localization process*, and can be viewed as a Markov chain on the space of probability measures. Localization processes based on tilts is important, as studying their associated Markov chains (see Lemma 2.1) can be reduced to studying the covariance matrices of the measures encountered during the localization process.

To keep the body of this lecture short, we give the SDE formulation of the coordinate-by-coordinate and negative fields localizations in Appendix B.

Example 7 (Gaussian Channel (Cont.)). Continuing Examples 2 and 5, let μ be a probability measure on $\Omega \subseteq \mathbb{R}^n$, and let $(B_t)_{t=0}^{\infty}$ be a Brownian motion on $\mathbb{R}^n$ whose randomness is independent of μ. We define $\mu^{(t+\,dt)}$ from $\mu^{(t)}$ by

$$\frac{\mu^{(t+\,dt)}(x)}{\mu^{(t)}(x)} \stackrel{\mathsf{def}}{=} 1 + \left\langle x - \boldsymbol{m}\left(\mu^{(t)}\right), A^{-1/2}\,dB_t \right\rangle \tag{4}$$

This recovers the Gaussian channel localization scheme via direct computation using Itô calculus. To get a rough sense on why these are equivalent, note that by rearranging Eq. (4) and applying *Itô's formula* for differentiating a function involving a Brownian motion term, we obtain

$$d\log \mu^{(t)}(x) = \left\langle x - \boldsymbol{m}\left(\mu^{(t)}\right), A^{-1/2}\,dB_t \right\rangle - \frac{1}{2}\left(x - \boldsymbol{m}\left(\mu^{(t)}\right)\right)^{\top} A^{-1}\left(x - \boldsymbol{m}\left(\mu^{(t)}\right)\right)\,dt. \tag{5}$$

If we drop the $\boldsymbol{m}\left(\mu^{(t)}\right)$ terms, which are present only to normalize things, we can then integrate both sides and rearrange to recover Eq. (1). Note that the first term in the right-hand side is what we would get if we used the usual rules of calculus, and is responsible for the external field component $\exp\left(x^{\top}Az_t\right)$ in Eq. (1). The second term is due to the randomness of dB_t, and accounts for the Gaussian component $\exp\left(-\frac{1}{2}tx^{\top}Ax\right)$ in Eq. (1). We have the second term because the standard deviation of dB_t is of order $\sqrt{dt}$, not dt, and so second-order/quadratic effects (e.g. from the Taylor expansion) remain macroscopic relative to dt, and must be included when we do stochastic differentiation.

2 Markov Chains via Localization

Every localization process gives rise to a natural Markov chain which we can use to sample from μ.

Lemma 2.1. *Let μ be a probability measure on a state space Ω, and let $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ be a localization process for μ. Then for every $t > 0$,[7], the family of probability measures $\{\mathsf{P}(x \to \cdot)\}_{x\in\Omega}$ on Ω given by*

$$\mathsf{P}(x \to A) \stackrel{\mathsf{def}}{=} \mathbb{E}\left[\frac{\mu^{(t)}(x)\mu^{(t)}(A)}{\mu(x)}\right], \qquad \forall x \in \Omega, A \subseteq \Omega, \tag{6}$$

is a Markov chain on Ω which is reversible w.r.t. μ.[8] Here, the expectation is again w.r.t. the randomness in the choice of measure $\mu^{(t)}$ (i.e. the mixture measure $\xi^{(t)}$).

Given the formula in Eq. (6), verifying that this is indeed a reversible Markov chain is straightforward. One way to express the action of this Markov chain is as follows: Suppose the current state is x_{prev}.

1. We first sample a random $\iota \in \mathcal{I}^{(t)}$ according to $\xi^{(t)}(\iota) \cdot \frac{\nu_\iota(x_{\mathsf{prev}})}{\mu(x_{\mathsf{prev}})}$, which indexes a probability measure ν_ι over Ω.
2. We then transition to $x_{\mathsf{next}} \sim \nu_\iota$.

[7] t is also allowed to be a random *stopping time*.

[8] We technically should write $\mathsf{P}^{(t)}$ instead of P to highlight the dependence of the Markov chain on t. However, this could be confused with iterating the Markov chain t times, so we suppress this notation.

To formalize this, let $\mathcal{D}^{(t)} \in \mathbb{R}_{\geq 0}^{\Omega \times \mathcal{I}^{(t)}}$ be the Markov kernel capturing the first step (i.e. mapping a random $x_{\mathsf{prev}} \sim \mu'$ for some other distribution μ' on Ω, to a random $\iota \in \mathcal{I}^{(t)}$). This $\mathcal{D}^{(t)}$ is precisely the noisy channel in Section 1.2. Similarly, let $\mathcal{U}^{(t)} \in \mathbb{R}_{\geq 0}^{\mathcal{I}^{(t)} \times \Omega}$ be the Markov kernel capturing the second step (i.e. mapping a random $\iota \sim \xi'$ for some other distribution ξ' on $\mathcal{I}^{(t)}$, to a random $x_{\mathsf{next}} \sim \nu_\iota$). Its rows are precisely the component measures $\left\{\nu_\iota : \iota \in \mathcal{I}^{(t)}\right\}$. This $\mathcal{U}^{(t)}$ precisely gives the conditional measure described in the statistical inference perspective (see Section 1.2). For the coordinate-by-coordinate localization process, $\mathcal{D}^{(t)}, \mathcal{U}^{(t)}$ are precisely the "down" and "up" operators in the spectral/entropic independence framework.

Fact 2.2 ("Inference Version" of Lemma 2.1)**.** *For* P *as defined in Lemma 2.1, we have that* $\mathsf{P} = \mathcal{D}^{(t)}\mathcal{U}^{(t)}$*. Furthermore,* $\mathcal{D}^{(t)}, \mathcal{U}^{(t)}$ *are* adjoints *of each other in the sense that for every pair of functions* $f : \mathcal{I}^{(t)} \to \mathbb{R}$ *and* $g : \Omega \to \mathbb{R}$*,*

$$\left\langle f, \mathcal{U}^{(t)} g \right\rangle_{\xi^{(t)}} = \left\langle \mathcal{D}^{(t)} f, g \right\rangle_{\mu}. \tag{7}$$

Another way to think about this Markov chain is via coupling: We draw a coupled pair of random variables (X, Y), each of which is marginally distributed according to μ, by

1. first sampling the random choice of (component) measure $\mu^{(t)}$ from the mixture measure $\xi^{(t)}$,
2. then sampling X, Y from $\mu^{(t)}$ independently from each other.

X and Y are coupled through the choice of $\mu^{(t)}$. The Markov chain described in Eq. (6) is the law of Y conditioned on knowing X. Let us now see some examples.

Example 8 (Glauber Dynamics). For the coordinate-by-coordinate localization (seer Examples 1, 4 and 11), Eq. (6) with $t = n - 1$ gives rise precisely to Glauber dynamics. Indeed, following the above coupling interpretation of P, we sample a coupled pair (σ, σ'), each marginally distributed according to μ, such that they agree on all except a single uniformly random coordinate. The conditional measure of σ' given σ is supported on configurations obtained from σ by flipping at most one coordinate. A direct calculation reveals that

$$\mathsf{P}\left(\sigma \to \sigma^{\oplus i}\right) = \frac{1}{n} \cdot \frac{\mu\left(\sigma^{\oplus i}\right)}{\mu(\sigma) + \mu\left(\sigma^{\oplus i}\right)}, \qquad \forall \sigma \in \{\pm 1\}^n, i \in [n].$$

For other choices of t, we also recover the *t-uniform block dynamics*, where a uniformly random $(n-t)$-subset of coordinates are resampled in each step.

Example 9 (Restricted Gaussian Dynamics). For the Gaussian channel localization (see Examples 2, 5 and 7), Eq. (6) for fixed finite $t \in \mathbb{R}_{\geq 0}$ gives rise to a Markov chain called the *restricted Gaussian dynamics*, first introduced in [LST21] in the context of continuous log-concave sampling. Let μ be a probability measure on $\Omega \subseteq \mathbb{R}^n$ (e.g. $\{\pm 1\}^n$, a convex body, all of $\mathbb{R}^n$, etc.). For a fixed $t \in \mathbb{R}_{\geq 0}$, its evolution proceeds as follows. Suppose we are currently at a configuration $x_{\mathsf{prev}} \in \Omega$.

- Draw $z_t \sim \mathcal{N}\left(t x_{\mathsf{prev}}, t A^{-1}\right)$.
- Sample x_{next} from the *Gaussian tilt* $\propto \mu(x_{\mathsf{next}}) \cdot \exp\left(-\frac{1}{2}\left(z_t - t x_{\mathsf{prev}}\right)^\top \frac{A}{t}\left(z_t - t x_{\mathsf{prev}}\right)\right)$ and transition to x_{next}.

It turns out that in the $t \to \infty$ limit, we also recover the classical *Langevin dynamics*.

Example 10 (Field Dynamics). For the negative fields localization (see Examples 3, 6 and 12), Eq. (6) gives rise to a Markov chain called the *field dynamics*, first introduced in [Che+21]. For a fixed $t \in \mathbb{R}_{\geq 0}$, its evolution proceeds as follows. Suppose we are currently at a configuration $\sigma \in \{\pm 1\}^n$.

- Draw a random subset of coordinates $S \subseteq \sigma^{-1}(+1)$ as follows: Every coordinate $i \in \sigma^{-1}(+1)$ is added independently with probability $\frac{1}{1+e^{-t}}$.
- Resample $\sigma_{[n]\setminus S}$ according to the distribution $(\mathcal{T}_{-t\mathbf{1}}\mu)^{S \leftarrow +\mathbf{1}}$.

The connection between the negative fields localization and the field dynamics was stated in [CE22] without proof, and written down more explicitly in [CLY23].

3 Conservation of Variance and Entropy

Now let us study how to establish fast mixing of the Markov chain P described in Lemma 2.1. We will significantly generalize the framework of spectral and entropic independence. First, observe that the Dirichlet form of P is given by[9]

$$\mathcal{E}_{\mathsf{P}}(f,g) = \mathbb{E}\left[\mathrm{Cov}_{\mu^{(t)}}(f,g)\right].$$

Hence, we can rewrite global variance and entropy factorization (i.e. Poincaré and modified/standard log-Sobolev constants) in the language of localization schemes as

$$\begin{aligned} \gamma(\mathsf{P}) &= \inf_{f:\Omega\to\mathbb{R}} \frac{\mathbb{E}\left[\mathrm{Var}_{\mu^{(t)}}(f)\right]}{\mathrm{Var}_{\mu}(f)} \\ \varrho(\mathsf{P}) &\geq \inf_{f:\Omega\to\mathbb{R}_{\geq 0}} \frac{\mathbb{E}\left[\mathrm{Ent}_{\mu^{(t)}}(f)\right]}{\mathrm{Ent}_{\mu}(f)} \end{aligned} \tag{8}$$

The right-hand side of each of these are essentially the optimal constants for "approximate tensorization of variance/entropy". These are global quantities which control the mixing time of P. Now, let us reduce these to more "local" inequalities.

Definition 2 ((Approximate) Conservation of Variance/Entropy; Informal)**.** *Let μ be a probability measure on a state space Ω, and let $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ be a localization process for μ. We say $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ satisfies* approximate conservation of variance *w.r.t a sequence of nonnegative constants $\left\{\theta^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ if for every global test function $f:\Omega\to\mathbb{R}$,*

$$\frac{\mathbb{E}\left[\mathrm{Var}_{\mu^{(t+\,dt)}}(f)\mid\mu^{(t)}\right]}{\mathrm{Var}_{\mu^{(t)}}(f)} \geq 1-\theta^{(t)}\,dt \tag{9}$$

We say $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ satisfies approximate conservation of entropy *if the same inequality, with all occurrences of* Var *replaced by* Ent*, holds for all global nonnegative test functions $f:\Omega\to\mathbb{R}_{\geq 0}$.*

Remark 2. This definition more or less makes sense for continuous-time localization processes. The discrete-time analog is simpler, since one should take the time increment to be $dt=1$.

Theorem 3.1 (Local-to-Global; Informal)**.** *Let μ be a probability measure on a state space Ω, and let $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ be a (continuous-time) localization process for μ. If $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ satisfies approximate conservation of variance (resp. entropy) w.r.t. $\left\{\theta^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$, then $\gamma(\mathsf{P}) \geq \exp\left(-\int_0^t \theta^{(s)}\,ds\right)$ (resp. $\varrho(\mathsf{P}) \geq \exp\left(-\int_0^t \theta^{(s)}\,ds\right)$).*

An informal proof is briefly sketched in Appendix C, although the essential idea is similar to what we already saw previously.

Relation with Spectral and Entropic Independence In the above framework, approximate conservation of entropy for the coordinate-by-coordinate localization process (see Examples 1, 4 and 11) is precisely entropic independence. In a similar manner, conservation of variance corresponds to spectral independence, although an additional step is required to show that it is equivalent to bounding the maximum eigenvalue of the influence matrix. This calculation was already done in a previous lecture. Theorem 3.1 and its proof is essentially the same as the one we used previously to show that spectral/entropic independence implies rapid mixing of Glauber dynamics. The crucial feature which makes this possible is that these are all *linear tilt* localization schemes.

3.1 Entropy Conservation via Correlation Inequalities

As we previously mentioned, approximate conservation of variance/entropy for the coordinate-by-coordinate localization scheme boils down to spectral/entropic independence. In a similar way, conservation for other localization schemes (e.g. Gaussian channel, negative fields, etc.) can also be reduced to correlation inequalities.

[9] Or, written more explicitly, $\mathbb{E}_{\iota\sim\xi^{(t)}}\left[\mathrm{Cov}_{\nu_\iota}(f,g)\right]$.

Lemma 3.2 (Entropy Conservation for Gaussian Channel Localization)**.** *Let μ be a probability measure on $\mathbb{R}^n$, and consider the Gaussian channel localization $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ of μ w.r.t. a driving matrix $A \succ 0$ (see Examples 2, 5 and 7). Fix some $t \in \mathbb{R}_{\geq 0}$, and suppose there is a sequence of positive semidefinite matrices $(K_s)_{s\in[0,t]}$ such that*

$$\mathrm{Cov}\left(\mathcal{T}_v \mu^{(s)}\right) \preceq K_s, \qquad \forall v \in \mathbb{R}^n, 0 \leq s \leq t \qquad \text{almost surely.}$$

Then we have the approximate conservation of entropy bound

$$\frac{\mathbb{E}\left[\mathrm{Ent}_{\mu^{(t)}}(f)\right]}{\mathrm{Ent}_{\mu}(f)} \geq \exp\left(-\|A\|_{\mathsf{op}}^{-1} \int_0^t \|K_s\|_{\mathsf{op}}\, ds\right).$$

Lemma 3.3 (Entropy Conservation for Negative Field Localization)**.** *Let μ be a probability measure on $\{\pm 1\}^n$, and consider the negative fields localization $\left\{\mu^{(t)}\right\}_{t\in\mathbb{R}_{\geq 0}}$ of μ (see Examples 3, 6 and 12). Suppose the following two conditions hold almost surely for $\mu^{(t)}$ for all $t \in \mathbb{R}_{\geq 0}$:*

- ***Spectral Independence:*** *There is a constant $\eta \geq 0$ (depending only on μ, not t) such that $\mu^{(t)}$ and all of its pinnings are all η-spectrally independent.*
- ***Tame/Weakly Stable Marginals:*** *There is a constant $\mathscr{K} > 1$ (depending only on μ, not t) such that $\mu^{(t)}$ and all of its pinnings satisfy*

$$\frac{\Pr[i \leftarrow +1 \mid \tau]}{\Pr[i \leftarrow -1 \mid \tau]} \leq \mathscr{K} \cdot \frac{\Pr[i \leftarrow +1]}{\Pr[i \leftarrow -1]}$$
$$\Pr[i \leftarrow -1] \geq \frac{1}{\mathscr{K}}.$$

Then there is a universal constant $c > 0$ such that for every $t \in \mathbb{R}_{\geq 0}$, we have the approximate conservation of entropy bound

$$\frac{\mathbb{E}\left[\mathrm{Ent}_{\mu^{(t)}}(f)\right]}{\mathrm{Ent}_{\mu}(f)} \geq \exp\left(-c\mathscr{K}^4 t\eta\right).$$

Remark 3. One can also replace all occurrences of entropy with variance in Lemma 3.2 and the statement will still hold. We can do the same for Lemma 3.3, in which case the "tame/weakly stable marginal" assumption is no longer required.

Remark 4. In Lemma 3.3, the "tame/weakly stable marginals" requirement is a relaxation of the "marginal boundedness" assumption we previously used.

Rather than formally proving Lemmas 3.2 and 3.3, which will require some background in stochastic calculus, let us give some heuristic intuition for their statements; we refer interested readers to [CE22] for all of the relevant calculations. For convenience, we look at conservation of variance rather than entropy; the case of entropy is handled similar to what we did previously, by either strengthening the correlation assumption to include all exponential tilts (as in Lemma 3.2) and then using the logarithmic Laplace transform, or imposing additional assumptions on the marginals (as in Lemma 3.3).

Intuition 1 (Lemmas 3.2 and 3.3). We argue along the lines of how we previously connected the influence matrix in spectral independence to factorization of variance (i.e. approximate conservation of variance for the coordinate-by-coordinate localization scheme). Fix some small time (increment) $dt \approx 0$, which has mixture measure $\xi^{(dt)}$ over $\mathcal{I}^{(dt)}$ and component measures $\left\{\nu_\iota : \iota \in \mathcal{I}^{(dt)}\right\}$. Then

$$\begin{aligned}
\inf_{f:\Omega\to\mathbb{R}} \frac{\mathbb{E}\left[\mathrm{Var}_{\mu^{(dt)}}(f)\right]}{\mathrm{Var}_{\mu^{(0)}}(f)} &= \gamma\left(\mathcal{D}^{(dt)}\mathcal{U}^{(dt)}\right) && \text{(Fact 2.2)} \\
&= \gamma\left(\mathcal{U}^{(dt)}\mathcal{D}^{(dt)}\right) && (AB \text{ and } BA \text{ have same (nonzero) eigenvalues}) \\
&= 1 - \lambda_{\max}\left(\mathcal{U}^{(dt)}\mathcal{D}^{(dt)} - \mathbf{1}\otimes\xi^{(dt)}\right).
\end{aligned}$$

Similar to $\mathcal{D}^{(dt)}\mathcal{U}^{(dt)}$, one can also view $\mathcal{U}^{(dt)}\mathcal{D}^{(dt)}$ through the lens of coupling: We draw a coupled pair (ι, ι'), each marginally distributed according to $\xi^{(dt)}$, by first sampling $x \sim \mu$, and then sampling $\iota, \iota' \sim \delta_x \mathcal{D}^{(dt)}$ independently. The pair (ι, ι') is correlated through the common $x \sim \mu$.

Because $dt \approx 0$ is small, one should think of $\mathcal{I}^{(dt)}$ as being a small set compared to the entire space $\Omega = \mathcal{I}^{(\infty)}$, and $\mathcal{D}^{(dt)}\mathcal{U}^{(dt)}$ being comparable to the trivial Markov chain $\mathcal{D}^{(0)}\mathcal{U}^{(0)} = \mathbf{1}\mu^\top$. For instance, in the extreme case $dt = 0$, then $\mathcal{I}^{(dt)}$ is a singleton, stemming from the fact that $\mu^{(0)} = \mu$ w.p. 1. As another example, if μ is on $\{\pm 1\}^n$, then for coordinate-by-coordinate localization with $dt = 1$, we have $\mathcal{I}^{(dt)} = [n] \times \{\pm 1\}$. One should then think of $\mathcal{U}^{(dt)}\mathcal{D}^{(dt)}$ as encoding "pairwise conditional marginals" of $\iota, \iota' \in \mathcal{I}^{(dt)}$ in μ, and $\mathcal{U}^{(dt)}\mathcal{D}^{(dt)} - \mathbf{1} \otimes \xi^{(dt)}$ as encoding "pairwise correlations".[10]

Now, here is the punchline for why the covariance matrix of μ itself appears in Lemmas 3.2 and 3.3: The coordinate-by-coordinate, Gaussian channel, and negative fields localization schemes are all *linear tilt* localization schemes! This means that these $\iota \in \mathcal{I}^{(dt)}$ can be encoded through vectors in $\mathbb{R}^n$, and their correlation structure is dictated by the inner product induced by the matrix $\mathrm{Cov}(\mu)$ (or that of various exponential/Gaussian tilts of μ). Hence, in essence, there is significant "dimensionality reduction"/"compression" when we ask for $\lambda_{\max}\left(\mathcal{U}^{(dt)}\mathcal{D}^{(dt)} - \mathbf{1} \otimes \xi^{(dt)}\right)$, which translates to a spectral condition on the covariance of μ and its tilts. The same intuition applies verbatim if we replace $\mu^{(0)} = \mu$ by $\mu^{(t)}$, and $\mu^{(dt)}, \mathcal{U}^{(dt)}, \mathcal{D}^{(dt)}$ by $\mu^{(t+dt)}, \mathcal{U}^{(t+dt)}, \mathcal{D}^{(t+dt)}$. This can all be formalized via stochastic calculus.

4 Annealing and Boosting

One of the most useful aspects of the localization schemes framework is the fact that one can concatenate localization schemes. Suppose one wishes to prove fast mixing of a natural Markov chain like Glauber dynamics. This naturally leads one to study the coordinate-by-coordinate localization per Example 8. However, in many settings, we do not know how to achieve optimal $O(n \log n)$ mixing with spectral/entropic independence for pinnings (i.e. coordinate-by-coordinate localization) alone. A major insight of [CE22] is to apply another "better behaved" localization scheme first, and then perform coordinate-by-coordinate localization to obtain tighter control on the mixing time of Glauber dynamics.

Definition 3 (Concatenation)**.** *Let $\mathcal{L}_{\mathsf{init}}, \mathcal{L}_{\mathsf{final}}$ be two localization schemes on a common state space Ω. For a (possibly random stopping) time $T \in \mathbb{R}_{\geq 0}$, define their T-*concatenation *as the localization scheme $\mathcal{L} = \mathsf{Concat}_T\left(\mathcal{L}_{\mathsf{init}}, \mathcal{L}_{\mathsf{final}}\right)$ which maps a measure μ on Ω to the following localization process:*

- *For every $0 \leq t \leq T$, let $\mu^{(t)} = \mathcal{L}_{\mathsf{init}}(\mu, t)$.*
- *For every $T < t < \infty$, let $\mu^{(t)} = \mathcal{L}_{\mathsf{final}}\left(\mu^{(T)}, t - T\right)$.*

Theorem 4.1 (Annealing/Boosting)**.** *Let $\mathcal{L}_{\mathsf{init}}, \mathcal{L}_{\mathsf{final}}$ be two localization schemes on a common state space Ω such that $\mathcal{L}_{\mathsf{final}}$ is* Doob *in the sense of* **??**. *For each probability measure ν on Ω, let P_ν denote the Markov chain obtained from $\mathcal{L}_{\mathsf{final}}$ via Eq. (6) for some specific time $t \in \mathbb{R}_{\geq 0}$.*

Now fix a probability measure μ on Ω, and let $\left\{\mu^{(t)}\right\}_{t \in \mathbb{R}_{\geq 0}}$ be the localization process produced by applying $\mathcal{L} = \mathsf{Concat}_T\left(\mathcal{L}_{\mathsf{init}}, \mathcal{L}_{\mathsf{final}}\right)$ to μ for some (possibly random stopping) time $T \in \mathbb{R}_{\geq 0}$. Assume the following two conditions hold:

- ***Variance/Entropy Conservation for $\mathcal{L}_{\mathsf{init}}$:*** *There exists a constant $\theta > 0$ such that*

$$\inf_{f:\Omega\to\mathbb{R}} \frac{\mathbb{E}\left[\mathrm{Var}_{\mu^{(T)}}(f)\right]}{\mathrm{Var}_\mu(f)} \geq \theta \qquad \left(\textit{resp.} \quad \inf_{f:\Omega\to\mathbb{R}_{\geq 0}} \frac{\mathbb{E}\left[\mathrm{Ent}_{\mu^{(T)}}(f)\right]}{\mathrm{Ent}_\mu(f)} \geq \theta\right).$$

- ***Fast Mixing of Components:*** *There exists a constant $\alpha > 0$ such that*

$$\gamma\left(\mathsf{P}_{\mu^{(T)}}\right) \geq \alpha \qquad \left(\textit{resp.}\ \varrho\left(\mathsf{P}_{\mu^{(T)}}\right) \geq \alpha\right) \qquad \textit{almost surely}$$

  (e.g. $\mathcal{L}_{\mathsf{final}}$ also satisfies approximate conservation of variance/entropy).

Then $\gamma\left(\mathsf{P}_\mu\right) \geq \theta\alpha$ (resp. $\varrho\left(\mathsf{P}_\mu\right) \geq \theta\alpha$).

[10] Again, in all of the above manipulations, there are technical issues which need to be addressed (e.g. the fact that $\mathcal{D}^{(t)}, \mathcal{U}^{(t)}$ are infinitely large matrices with possibly uncountably many rows, etc.), and this can all be resolved through the language of stochastic calculus.

Since the proof of Theorem 4.1 requires the notion of a *Doob localization scheme* (see **??**), we relegate its proof to Appendix A. It suffices to say that all of our prime examples, i.e. coordinate-by-coordinate (Example 4), Gaussian channel (Example 5), and negative fields (Example 6), are all Doob. The strength of Theorem 4.1 is that we can use $\mathcal{L}_{\mathsf{init}}$ to first *anneal* the distribution μ and make it more well-behaved, before applying $\mathcal{L}_{\mathsf{final}}$ which ultimately induces the Markov chain of interest. To illustrate this tool, let us establish optimal $O(n \log n)$-mixing of Glauber dynamics for two models which we do not know how to achieve via the tools we've developed so far (although polynomial mixing times can be achieved via spectral/entropic independence for pinnings).

4.1 Ising Models with Bounded Spectral Diameter

Theorem 4.2 ([EKZ22; Ana+22; CE22])**.** *Fix a constant $0 < \delta < 1$, and let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix satisfying $\lambda_{\max}(A) - \lambda_{\min}(A) \le 1 - \delta$. Then Glauber dynamics for the corresponding Ising Gibbs measure $\mu(\sigma) \propto \exp\left(\frac{1}{2}\sigma^\top A \sigma\right)$ over $\{\pm 1\}^n$ satisfies $\varrho(\mathsf{P}_{\mathsf{GD}}) \ge \frac{\delta}{n}$.*

Remark 5. A major result of Bauerschmidt–Bodineau [BB19], preceding [EKZ22; Ana+22; CE22], obtained an optimal log-Sobolev inequality with a slight modification of the Glauber Dirichlet form. This unfortunately did not imply fast mixing of Glauber dynamics, but it stimulated much further work.

Proof Sketch. Note that since μ is supported on $\{\pm 1\}^n$, it is invariant under shifting A by a diagonal matrix. In particular, we can assume $\frac{\delta}{2} \cdot \mathsf{Id} \preceq A \preceq \left(1 - \frac{\delta}{2}\right) \cdot \mathsf{Id}$. We use Gaussian channel localization with driving matrix $A^{-1/2}$ as $\mathcal{L}_{\mathsf{init}}$, and coordinate-by-coordinate localization as $\mathcal{L}_{\mathsf{final}}$. We concatenate at time $T = 1$, since at precisely this time, the Gaussian component has completely canceled out the original quadratic interaction term, leaving only a product measure for $\mu^{(t)}$.

We let us now formalize the argument. Throughout, P_μ is precisely Glauber dynamics for μ. We verify the conditions of Theorem 4.1. By the definition of $\mathcal{L}_{\mathsf{init}}$, we have that at the specially chosen time $T = 1$, $\mu^{(T)} = \mathcal{T}_v \mathsf{Unif}\{\pm 1\}^n$ for some $v \in \mathbb{R}^n$. Hence, $\varrho\left(\mathsf{P}_{\mu^{(T)}}\right) \ge 1/n$ almost surely.[11] We just need approximation conservation of entropy for $\mathcal{L}_{\mathsf{init}}$ with constant δ. This is obtained by verifying Lemma 3.2. Indeed, since $\mu^{(s)}(\sigma) \propto \exp\left(\frac{1-s}{2}\sigma^\top A \sigma + \langle \sigma, v \rangle\right)$ for some $v \in \mathbb{R}^n$ and every $0 \le s \le T = 1$, by the results of the third problem set, we have that $\mathrm{Cov}\left(\mathcal{T}_u \mu^{(s)}\right) \preceq \frac{1}{1 - s(1-\delta)}$ almost surely for all $v \in \mathbb{R}^n$. Taking $K_s = \frac{1}{1 - s(1-\delta)} \cdot \mathsf{Id}$, Lemma 3.2 says that

$$\theta \ge \exp\left(-\int_0^1 \frac{1-\delta}{1 - s(1-\delta)}\, ds\right) = \delta.$$

Combined with $\varrho\left(\mathsf{P}_{\mu^{(T)}}\right) \ge 1/n$, Theorem 4.1 implies that $\varrho(\mathsf{P}_\mu) \ge \frac{\delta}{n}$ as desired. □

4.2 Tree-Unique Hardcore Model without Bounded Maximum Degree

Theorem 4.3 ([Che+21; Ana+22; CE22])**.** *Let $G = (V, E)$ be an arbitrary graph, and let Δ denote its maximum degree (which we now allow to grow with n). Let $0 < \delta < 1$ be a fixed constant, and assume $\lambda \le (1 - \delta)\lambda_c(\Delta)$. Then there is a universal constant $c > 0$ such that Glauber dynamics for the hardcore Gibbs measure on G with activity λ satisfies $\varrho(\mathsf{P}_{\mathsf{GD}}) \ge \frac{\exp(-c/\delta)}{n}$.*

Proof Sketch. To prove Theorem 4.3, we use the negative fields localization as $\mathcal{L}_{\mathsf{init}}$, and coordinate-by-coordinate localization as $\mathcal{L}_{\mathsf{final}}$. We concatenate at a constant $T \approx \log(100e)$. This time is chosen so that almost surely, the measure $\mu^{(T)}$ is the hardcore Gibbs measure of some subgraph of G with activity $e^{-T}\lambda \approx \frac{1}{100\Delta}$. For this, it was previously known that $\varrho\left(\mathsf{P}_{\mu^{(T)}}\right) \gtrsim \frac{1}{n}$ up to a multiplicative loss by a universal constant [Erb+17].[12] To apply Theorem 4.1, all that remains is to establish approximate conservation of entropy along the negative fields localization for μ.

For this, observe that $\mu^{(s)}$ is always a hardcore Gibbs measure on a subgraph of G with a weaker activity $e^{-s}\lambda$. Hence, by the "correlation decay to spectral independence" results we have already seen, we have $O(1/\delta)$-spectral independence for $\mu^{(s)}$ and all of its pinnings. The "tame/weakly stable marginals" condition can also be verified directly for the hardcore model, using the fact

[11] We more or less already proved this by tensorization of entropy earlier when we discussed spectral/entropic independence.

[12] Note that this is well within the regime in which path coupling contracts. However, an $\Omega(1/n)$ lower bound on the modified log-Sobolev constant is still highly nontrivial.

that the marginal lower and upper bounds coincide up to a universal multiplicative constant when $\lambda \leq O(1/\Delta)$. We omit the calculation for brevity. Applying Lemma 3.3, we obtain the desired cnoservation of entropy. □

References

[AM22] Ahmed El Alaoui and Andrea Montanari. "An Information-Theoretic View of Stochastic Localization". In: *IEEE Transactions on Information Theory* 68.11 (2022) (cit. on p. 4).

[AMS22] A. Alaoui, A. Montanari, and M. Sellke. "Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. Los Alamitos, CA, USA: IEEE Computer Society, Nov. 2022, pp. 323–334. DOI: 10.1109/FOCS54457.2022.00038 (cit. on p. 1).

[Ana+22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Entropic Independence: Optimal Mixing of down-up Random Walks". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 1418–1430 (cit. on p. 10).

[BB19] Roland Bauerschmidt and Thierry Bodineau. "A very simple proof of the LSI for high temperature spin systems". In: *Journal of Functional Analysis* 276.8 (2019), pp. 2582–2588. ISSN: 0022-1236. DOI: https://doi.org/10.1016/j.jfa.2019.01.007 (cit. on p. 10).

[BBD23] Roland Bauerschmidt, Thierry Bodineau, and Benoit Dagallier. "Stochastic dynamics and the Polchinski equation: an introduction". In: *arXiv preprint arXiv:2307.07619* (2023) (cit. on p. 1).

[CE22] Yuansi Chen and Ronen Eldan. "Localization Schemes: A Framework for Proving Mixing Bounds for Markov Chains (extended abstract)". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. Los Alamitos, CA, USA: IEEE Computer Society, Nov. 2022, pp. 110–122. DOI: 10.1109/FOCS54457.2022.00018 (cit. on pp. 1, 2, 6, 8–10).

[Che+21] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. "Optimal Mixing Time for the Ising Model in the Uniqueness Regime". In: *arXiv preprint arXiv:2111.03034* (2021) (cit. on pp. 6, 10).

[Che21] Yuansi Chen. "An Almost Constant Lower Bound of the Isoperimetric Coefficient in the KLS Conjecture". In: *Geometric and Functional Analysis* 31 (2021), pp. 34–61 (cit. on p. 1).

[CLY23] Xiaoyu Chen, Jingcheng Liu, and Yitong Yin. "Uniqueness and Rapid Mixing in the Bipartite Hardcore Model". In: *arXiv preprint arXiv:2305.00186* (2023) (cit. on p. 6).

[EKZ22] 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* 182 (2022), pp. 1035–1051 (cit. on pp. 5, 10).

[Eld13] Ronen Eldan. "Thin Shell Implies Spectral Gap Up to Polylog via a Stochastic Localization Scheme". In: *Geometric and Functional Analysis* 23 (2013), pp. 532–569 (cit. on pp. 1, 2, 5).

[Erb+17] Matthias Erbar, Christopher Henderson, Georg Menz, and Prasad Tetali. "Ricci curvature bounds for weakly interacting Markov chains". In: *Electronic Journal of Probability* 22 (2017), pp. 1–23 (cit. on p. 10).

[KLS95] Ravi Kannan, László Lovász, and Miklós Simonovits. "Isoperimetric problems for convex bodies and a localization lemma". In: *Discrete & Computational Geometry* 13 (1995), pp. 541–559 (cit. on p. 1).

[LST21] Yin Tat Lee, Ruoqi Shen, and Kevin Tian. "Structured Logconcave Sampling with a Restricted Gaussian Oracle". In: *Proceedings of Thirty Fourth Conference on Learning Theory*. Ed. by Mikhail Belkin and Samory Kpotufe. Vol. 134. Proceedings of Machine Learning Research. PMLR, 15–19 Aug 2021, pp. 2993–3050 (cit. on p. 6).

[LV17] Yin Tat Lee and Santosh Srinivas Vempala. "Eldan's Stochastic Localization and the KLS Hyperplane Conjecture: An Improved Lower Bound for Expansion". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. 2017, pp. 998–1007. DOI: 10.1109/FOCS.2017.96 (cit. on p. 1).

[LV18] Yin Tat Lee and Santosh S. Vempala. "Stochastic Localization + Stieltjes Barrier = Tight Bound for Log-Sobolev". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 1122–1129. ISBN: 9781450355599. DOI: 10.1145/3188745.3188866 (cit. on p. 1).

[MW23] Andrea Montanari and Yuchen Wu. "Posterior Sampling from the Spiked Models via Diffusion Processes". In: *arXiv preprint arXiv:2304.11449* (2023) (cit. on p. 1).

A Filtrations and Doob Localization Schemes

As we said in Remark 1, the informal description provided at the beginning of Section 1.2 is inadequate when it comes to doing things formally. Here, we fill in the gap, as well as prove Theorem 4.1. We begin with some standard measure theory. TODO

Definition 4 (Filtration)**.** *TODO*

Proof of Theorem 4.1. TODO □

B Coordinate-by-Coordinate and Negative Fields as Linear Tilt Localizations

Example 11 (Coordinate-by-Coordinate (Cont.))*.* Continuing Examples 1 and 4, let μ be a probability measure over $\{\pm 1\}^n$, and let $(i_1, \dots, i_n)$ be a uniformly random permutation of $[n]$ whose randomness is independent of μ. Let $U_1, \dots, U_n$ be i.i.d. $\mathsf{Unif}[-1, 1]$ random variables, and for all $j \in [n]$, define

$$v_j \stackrel{\text{def}}{=} \frac{s_j}{1 + s_j \cdot \boldsymbol{m}\left(\mu^{(j-1)}\right)_{i_j}} \cdot \mathbf{1}_{i_j} \qquad \text{where} \qquad s_j = \operatorname{sign}\left(\boldsymbol{m}\left(\mu^{(j-1)}\right)_{i_j} - U_j\right).$$

Then take

$$\frac{\mu^{(t+1)}(\sigma)}{\mu^{(t)}(\sigma)} \stackrel{\text{def}}{=} 1 + \left\langle \sigma - \boldsymbol{m}\left(\mu^{(t)}\right), v_{t+1} \right\rangle, \qquad \forall \sigma \in \{\pm 1\}^n.$$

This again recovers the coordinate-by-coordinate localization scheme, since in each increment of t, the linear function $\left\langle \sigma - \boldsymbol{m}\left(\mu^{(t)}\right), v_{t+1} \right\rangle$ restricts $\mu^{(t)}$ is a subcube.

Example 12 (Negative Fields (Cont.))*.* Continuing Examples 3 and 6, let μ be a probability measure over $\{\pm 1\}^n$. We define $\mu^{(t+\,dt)}$ from $\mu^{(t)}$ by

$$\frac{\mu^{(t+1)}(\sigma)}{\mu^{(t)}(\sigma)} \stackrel{\text{def}}{=} 1 + \left\langle \sigma - \boldsymbol{m}\left(\mu^{(t)}\right), dv_t \right\rangle, \qquad \forall \sigma \in \{\pm 1\}^n,$$

where

$$v_t(i) \stackrel{\text{def}}{=} -t + \frac{\mathbf{1}[i \in S_{t+\,dt} \setminus S_t]}{1 + \boldsymbol{m}\left(\mu^{(t)}\right)_i}, \qquad \forall i \in [n],$$

and the S_t is built up as in Example 6.

C Unfinished Proofs

Proof Sketch of Theorem 3.1. Rearranging Eq. (9), we obtain that

$$d \log \mathbb{E}\left[\operatorname{Var}_{\mu^{(t)}}(f)\right] \geq -\theta^{(t)}\, dt.$$

Integrating from 0 to t,

$$\log \mathbb{E}\left[\mathrm{Var}_{\mu^{(t)}}(f)\right] - \log \mathrm{Var}_{\mu}(f) \geq -\int_0^t \theta^{(s)}\, ds.$$

Exponentiating both sides and using Eq. (8) yields the claim. □

Remark 6. The manipulations here don't quite make sense for discrete-time localization processes. However, the analog is to replace the integral with a sum, and use the chain rule plus a telescoping trick:

$$\begin{aligned}
\frac{\mathbb{E}\left[\mathrm{Var}_{\mu^{(t)}}(f)\right]}{\mathrm{Var}_{\mu}(f)} &= \mathbb{E}\left[\prod_{s=1}^{t} \frac{\mathrm{Var}_{\mu^{(s)}}(f)}{\mathrm{Var}_{\mu^{(s-1)}}(f)}\right] \\
&= \mathbb{E}\left[\prod_{s=1}^{t} \mathbb{E}\left[\frac{\mathrm{Var}_{\mu^{(s)}}(f)}{\mathrm{Var}_{\mu^{(s-1)}}(f)} \,\middle|\, \mu^{(s-1)}\right]\right] \\
&\geq \prod_{s=1}^{t} \left(1 - \theta^{(s)}\right) \\
&\gtrsim \exp\left(-\sum_{s=1}^{t} \theta^{(s)}\right).
\end{aligned}$$