# Localization schemes: A framework for proving mixing bounds for Markov chains

Yuansi Chen* and Ronen Eldan†

**Abstract**

Two recent and seemingly-unrelated techniques for proving mixing bounds for Markov chains are: (i) the framework of Spectral Independence, introduced by Anari, Liu and Oveis Gharan, and its numerous extensions, which have given rise to several breakthroughs in the analysis of mixing times of discrete Markov chains and (ii) the Stochastic Localization technique which has proven useful in establishing mixing and expansion bounds for both log-concave measures and for measures on the discrete hypercube. In this paper, we introduce a framework which connects ideas from both techniques. Our framework unifies, simplifies and extends those two techniques. In its center is the concept of a "localization scheme" which, to every probability measure on some space $\Omega$, assigns a martingale of probability measures which "localize" in space as time evolves. As it turns out, to every such scheme corresponds a Markov chain, and many chains of interest appear naturally in this framework. This viewpoint provides tools for deriving mixing bounds for the dynamics through the analysis of the corresponding localization process. Generalizations of concepts of Spectral Independence and Entropic Independence naturally arise from our definitions, and in particular we recover the main theorems in the spectral and entropic independence frameworks via simple martingale arguments (completely bypassing the need to use the theory of high-dimensional expanders). We demonstrate the strength of our proposed machinery by giving short and (arguably) simpler proofs to many mixing bounds in the recent literature. In particular, we: (i) Give the first $O(n \log n)$ bound for mixing time of the hardcore-model (of arbitrary degree) in the tree-uniqueness regime, under Glauber dynamics, (ii) Give the first optimal mixing bounds for Ising models in the uniqueness regime under any external fields, (iii) Prove a KL-divergence decay bound for log-concave sampling via the Restricted Gaussian Oracle, which achieves optimal mixing under any $\exp(n)$-warm start, (iv) Prove a logarithmic-Sobolev inequality for near-critical Ferromagnetic Ising models, recovering in a simple way a variant of a recent result by Bauerschmidt and Dagallier.

*Duke University.

†Microsoft Reseach. This work was partially supported by a European Research Council Grant no. 803084. and by NSF grant no. DMS-1926686.

# Contents

# 1 Introduction

Suppose that we would like to sample from a measure $\nu$ on some set $\Omega$. For the sake of discussion, suppose that either $\Omega = \{-1,1\}^n$ is the Boolean hypercube or $\Omega = \mathbb{R}^n$. A common algorithm is to find a Markov chain whose stationary distribution is $\nu$ and which exhibits good *mixing bounds*.

In the case $\Omega = \{-1,1\}^n$, a very useful Markov chain associated to a measure $\nu$ is the Glauber dynamics, defined as follows: Given $x \in \{-1,1\}^n$, the transition kernel from $x$, $P_{x\to}\cdot$, is the law which describes the point $Y$ picked according to following random procedure: Pick uniformly a coordinate $i \in [n]$ and then take $Y$ according to the law $\nu$ conditioned on the event $\{Y_j = x_j,\ \forall j \in [n] \setminus \{i\}\}$.

For a Markov chain $(X_t)_t$ in a state space $\Omega$ which has a unique stationary measure $\nu$, a mixing bound typically asserts that for every $\varepsilon > 0$ there is a time $t(\varepsilon)$ such that for all measurable $A \subset \Omega$ and all $t > t(\varepsilon)$, one has $|\mathbb{P}(X_t \in A) - \nu(A)| \leq \varepsilon$. See [LP17] for an extensive account of this subject.

In recent years, there appeared two seemingly-unrelated new techniques which were used to establish mixing bounds through functional inequalities:

- The work [ALOG20] put forth the notion of *Spectral independence* and developed a framework which relies on those notions in order to establish mixing bounds for measures on the set of subsets of $[n]$. This framework relies on the theory of high-dimensional expanders. Some follow-up works which extended this technique are [AJK+21b, CLV21, AJK+21a, CFYZ21b, FGYZ21, BCC+22, Liu21, AASV21].

- The *stochastic localization* technique, introduced by the second author in [Eld13], is the central ingredient used in the proofs of several functional inequalities, both in the continuous setting where $\Omega = \mathbb{R}^n$ and $\nu$ is a *logarithmically-concave* measure and in the setting of the discrete hypercube. Most notably, the technique gives the state-of-the-art bounds, due to the first author ([Che21]), for the so-called Kannan-Lovász-Simonovits conjecture ([KLS95a]) and Bourgain's slicing problem (see [KM21]). Some follow-up works based on this technique are [Eld20, ES22, Kla18, LV17, Che21].

In this work, we both unify and expand these two techniques towards a new framework which can be used to establish mixing bounds in various settings, showing that the same principles govern in a wide variety scenarios.

One of the main principles underlying both techniques is that concentration bounds on a measure can be deduced from bounds on the covariance structure of a certain family of measures which are transformations of the original measure: In the spectral independence framework, a

sufficient condition for a spectral gap is the boundedness of the *influence* matrices of restrictions of the measure (the influence matrix has a simple correspondence with the covariance matrix), and in the stochastic localization framework a spectral gap is implied by the boundedness of the covariance matrix along a certain stochastic process which is associated with the measure.

This work shows that with the correct point of view, those two reductions follow from the exact same argument. By approaching the notions of spectral and entropic independence from this point of view, we will be able to:

(i) Generalize the theory, giving rise to a natural family of Markov chains together with a toolbox of ingredients that can be used to prove mixing bounds for those chains.

(ii) Simplify the proofs in the foundations of the framework of spectral and entropic independence, and in particular completely bypass the need to use the theory of high-dimensional expanders.

(iii) Provide a self-contained and (arguably) simpler proofs for many of the expansions of the spectral/entropic independence machinery, and in particular reprove in a more general context several main theorems which appear in [ALOG20, CLV21, AJK+21b, AJK+21a, CFYZ21b, CFYZ21a, STL20, LST21].

**Summary of applications**

To demonstrate the strength of our machinery, we apply it in several settings:

- We derive the optimal mixing rate for Glauber dynamics on the hardcore model of any degree in the tree-uniqueness regime, showing that it mixes in time $O(n \log n)$. A similar mixing rate was obtained in [AJK+21a] for a different Markov chain which was tailored for the hardcore model. Additionally, our framework (arguably) allows us to significantly simplify the argument.

- We give the first optimal mixing bounds for graphical Ising models in the uniqueness regime under any external fields, improving the results in [CFYZ21a] in the sense that there is no dependence of the bound on the external field. In this case as well, the proof is significantly simpler.

- Provide a very simple proof of a KL-divergence decay bound for log-concave sampling via the so-called Restricted-Gaussian-Oracle introduced in [STL20]. Our bound works under any $\exp(n)$-warm start, resolving an issue raised in [LST21].

- Give a self-contained and simpler proof of mixing for the Glauber dynamics for Ising models whose interaction matrix has operator norm bounded by $1$, which in particular gives optimal mixing for Glauber dynamics on the Sherrington-Kirkpatrick model in high enough temperature, recovering the result derived from [EKZ21, AJK+21a].

- We recover, in a simple way, a variant of a recent result by Bauerschmidt and Dagallier [BD22], proving a logarithmic-Sobolev inequality for Ferromagnetic Ising models in terms of the model's susceptibility.

### Concurrent work

Shortly before submitting this manuscript, we were informed by Chen, Feng, Yin and Zhang of a manuscript in preparation [CFYZ22] which independently proves parts of the results regarding antiferromagnetic Ising models that we prove here, and in particular gives the same mixing bound for the hardcore model, using a different proof technique.

### Mixing via functional inequalities

Consider a reversible transition operator $P = P_{x\to y}$ with stationary measure $\nu$ on a state space $\Omega$. Its **spectral gap** is defined as the quantity

$$\mathfrak{gap}(P) := \inf_{\varphi:\Omega\to\mathbb{R}} \frac{\int_{\Omega\times\Omega}(\varphi(x)-\varphi(y))^2 dP_x(y)d\nu(x)}{2\mathrm{Var}_\nu[\varphi]}.$$

Next, define

$$\mathrm{Ent}_\nu[f] = \int f(x)\log f(x) d\nu(x) - \int f(x)d\nu(x)\log\left(\int f(x)d\nu(x)\right).$$

We define the **modified log-Sobolev Inequality (MLSI)** coefficient as

$$\rho_{\mathrm{LS}}(P) := 1 - \sup_{f:\Omega\to[0,\infty)} \frac{\mathrm{Ent}_\nu[Pf]}{\mathrm{Ent}_\nu[f]}.$$

*Remark* 1. The above definition is only valid if the Markov transition kernel $P$ is reversible (or self adjoint). If it is not reversible, $Pf$ needs to be replaced by $\frac{dP^*(f\nu)}{d\nu}$. In this work we only discuss the reversible case, so we stick to the simpler notation.

It is standard to deduce mixing bounds for the associated Markov chain from the above quantities. Given an initial distribution $\mu$ which is absolutely continuous with respect to $\nu$, consider the total-variation mixing time

$$t_{\mathrm{mix}}(P,\varepsilon;\mu) = \min\left\{t>0;\quad \left|P^t[\mu](A)-\nu(A)\right| \le \varepsilon,\ \ \forall A\subset\Omega\right\}$$

Moreover, define

$$t_{\mathrm{mix}}(P,\varepsilon) = \max_{x\in\Omega} t_{\mathrm{mix}}(P,\varepsilon,\delta_x).$$

The following fact is standard (see e.g. [LP17, Theorem 12.4] and [CLV21, Fact 3.5]).

**Fact 2.** *Suppose that for all $x\in\Omega$ one has $\nu(\{x\})\ge\eta$. Then,*

$$t_{\mathrm{mix}}(P,\varepsilon) \le C\mathfrak{gap}(P)^{-1}\left(\log(1/\eta)+\log(1/\varepsilon)\right), \text{ and}$$
$$t_{\mathrm{mix}}(P,\varepsilon) \le C\rho_{\mathrm{LS}}(P)^{-1}\left(\log\log(1/\eta)+\log(1/\varepsilon)\right).$$

The main aim of our machinery is to provide tools which give lower bounds for the spectral gap and modified log-Sobolev coefficient for a family of Markov chains that arise naturally in our framework.

## 1.1 Review of the framework and ideas

We now give a brief overview of the framework constructed in the next sections.

- The key definitions in the framework are a localization process and a localization scheme. A **localization process** is a stochastic process $(\nu_t)_{t\geq 0}$ where each $\nu_t$ is a probability measure on some space $\Omega$, having the property that for every subset $A \subset \Omega$, the stochastic process $t \to \nu_t(A)$ is a martingale which satisfies $\lim_{t\to\infty} \nu_t(A) \in \{0, 1\}$. A **localization scheme** is a mapping which, to every probability measure $\nu$ on $\Omega$, assigns a localization process $(\nu_t)_t$ with $\nu_0 = \nu$. Thus, a localization scheme can be thought of as a way to interpolate between a given measure $\nu$ and a (random) Dirac measure, via a martingale on the space of measures.

- Given a localization process $(\nu_t)_t$ and a time $\tau > 0$, there is a reversible Markov chain on $\nu$, whose stationary measure is $\nu = \nu_0$, which is naturally associated to the localization process. This Markov chain is defined by the formula

$$P_{x\to y} = \mathbb{E}\left[\frac{\nu_\tau(x)\nu_\tau(y)}{\nu(x)}\right].$$

  It turns out that many Markov chains arise via naturally defined localization schemes. In particular, the Glauber dynamics on $(\{-1, 1\}^n, \nu)$ can be derived from what we call the coordinate-by-coordinate localization scheme. This scheme is defined by taking $(k_1, ..., k_n)$ to be a uniformly random permutation of $[n]$, taking $X \sim \nu$ and setting $\nu_t$ to be the law of $X$ conditioned on $X_{k_1}, \ldots, X_{k_t}$. Some other chains that arise via this framework (using other localization schemes) are the hit-and-run walk, the up-down walk, the Restricted-Gaussian dynamics and the field dynamics.

- Next, we will see that there is a simple way to analyze the spectral gap and MLSI coefficient of the Markov chain associated to a localization scheme: In order to give a bound on the spectral gap, one needs to give a lower bound to the quantity $\mathbb{E}\left[\frac{\mathrm{Var}_{\nu_\tau}[\varphi]}{\mathrm{Var}_\nu[\varphi]}\right]$ for an arbitrary test function $\varphi : \Omega \to \mathbb{R}$. Similarly, for MLSI, one needs to lower bound the quantity $\mathbb{E}\left[\frac{\mathrm{Ent}_{\nu_\tau}[f]}{\mathrm{Ent}_\nu[f]}\right]$ for $f : \Omega \to [0, \infty)$. Intuitively speaking, the measure-valued process $\nu_t$ "zooms in" on smaller and smaller portions of $\Omega$, and we want to establish that there is still some variance (or entropy) left all the way up to time $\tau$, while the measure $\nu_\tau$ is already focused on a small portion of the space.

- In light of the above intuition, it makes sense to analyze the time differentials of the stochastic processes $t \to \log(\mathrm{Var}_{\nu_t}[\varphi])$ and $t \to \log(\mathrm{Ent}_{\nu_t}[f])$ (which can be in either continuous or discrete time, depending on the localization scheme). Roughly speaking, we want to obtain lower bounds for their drifts. By integrating those bounds with respect to time, we could then obtain bounds on the spectral gap and MLSI. When a measure $\nu$ satisfies the lower bound $\frac{d\mathrm{Var}_{\nu_t}[\varphi]}{\mathrm{Var}_{\nu_t}[\varphi]} \geq -\alpha dt +$ martingale, we say that it satisfies $\alpha$-**approximate conservation of variance** bound, and similarly when $\frac{d\mathrm{Ent}_{\nu_t}[f]}{\mathrm{Ent}_{\nu_t}[f]} \geq -\alpha dt +$ martingale, we say that it satisfies $\alpha$-**approximate conservation of entropy**.

- It turns out that there is an efficient way to obtain approximate conservation of variance and entropy bounds for a family of localization schemes which we call linear-tilt localizations. Roughly speaking, these are schemes where $\frac{\nu_{t+h}}{\nu_t}$ is a linear function up to

$o(h)$. We will see that many localization schemes of interest, including the coordinate-by-coordinate scheme and the stochastic localization scheme of [Eld13] can be described this way.

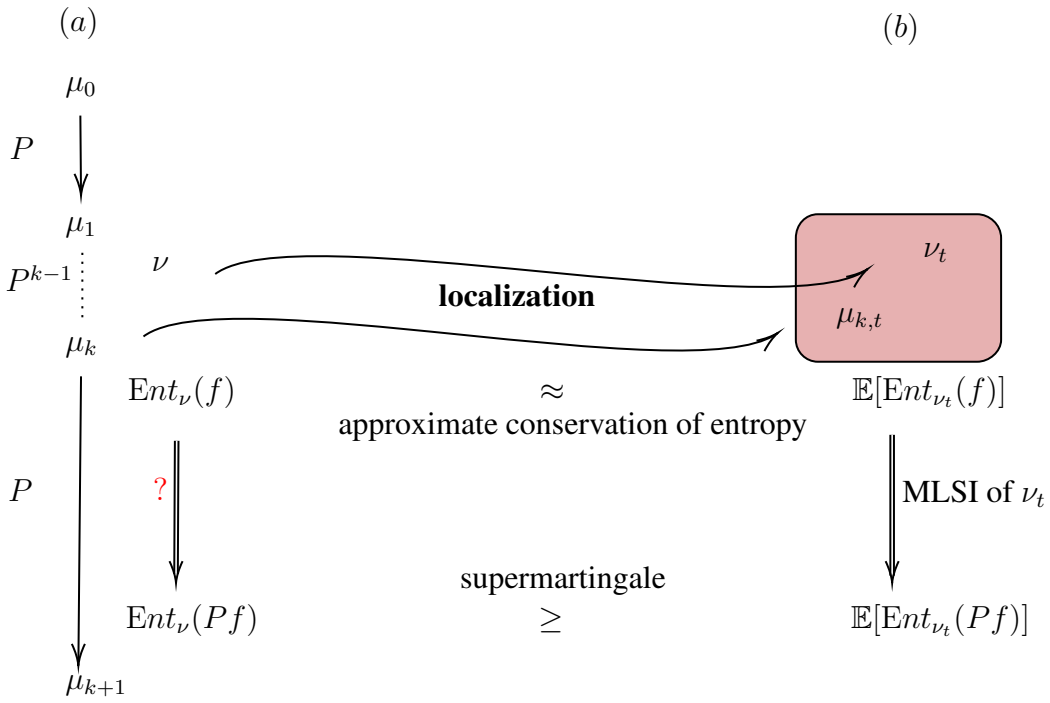


Figure 1: $(a)$. Probability measures encountered during the iterates of a Markov chain $P$ with target distribution $\nu$ and initial distribution $\mu_0$. We wish to establish the entropy decay (or MLSI) for each iteration of the Markov chain. $(b)$. Probability measures generated via a localization scheme. If a stochastic scheme is used, the red box indicates that one may generate multiple random instances of $(\nu_t, \mu_{k,t})$ starting from $(\nu,\ \mu_k)$. The localization scheme is designed such that it is easier to establish the entropy decay for the new measures on $(b)$. With $f = \frac{\mu_k}{\nu}$, the entropy decay properties of the measures on $(b)$ can be related to those on $(a)$ thanks to the martingale properties of the localization.

- For linear-tilt localization scheme, we see that a simple argument which uses the Cauchy-Schwartz inequality shows that approximate variance conservation is related to the process of covariance matrices $t \to \mathrm{Cov}(\nu_t)$. For the case of coordinate-by-coordinate localization this argument recovers the spectral independence framework of Anari, Liu and Oveis Gharan in a very simple way that completely bypasses the need to use high-dimensional expanders.

- Approximate conservation of entropy turns out to follow from a natural condition that we call **entropic stability**, which roughly says that the center of mass of the measure cannot move too much under perturbations with small relative entropy. This can be thought of as a weak form of **transportation-entropy inequalities**. In the context of the coordinate-by-coordinate localization we will see that it amounts to a very simple property of the logarithmic Laplace transform which is very similar to fractional log-concavity. This would give us a way to recover a variant of the entropic-independence framework introduced in [AJK+21b].

- Another ingredient in our framework will be the concept of **annealing** via a localization

scheme. Given two localization schemes on the same space, we can "concatenate" them by running one scheme up to some time $t$ and then running the other on the measure $\nu_t$ produced by the first scheme. What we will see is that we can produce spectral gap and MLSI bounds, with respect to the dynamics associated with the second localization scheme, by showing that the first localization scheme does not contract the variance / entropy and "anneals" the measure in a way that it outputs a measure to which we can ensure mixing for the second scheme. Such annealing procedure will be useful in many of the applications.

In Figure 1, we summarize the conceptual diagram used to analyze the MLSI coefficient of a Markov chain via a localization scheme.

## 1.2 Related techniques

### Local-to-global theorems and high-dimensional expanders

The framework of spectral independence was originally based on *local-spectral-expansion* inequalities in high-dimensional expanders, which was defined and derived in works of Alev, Dinur, Kaufmann, Lau, Mass and Oppenheim ([AL20, Opp18, KO20, DK17, KM17]). These inequalities allow us to compare between the spectral gap of the $\ell$-down-up-walk and that of the $\ell+1$-up-down-walk on spectral expanders. This principle was first used towards establishing bounds on the spectral gap by Anari, Liu, Oveis Gharan and Vinzant [ALGV19]. Our framework could be thought as an alternative way to prove "local-to-global" theorems.

### Pathwise analysis

Our technique can be seen as a manifestation of the pathwise analysis method (see [Eld22]). This method is closely-related to the so-called *semigroup technique* where, in order to prove an inequality regarding a measure, one considers an evolution of the measure via a semigroup (usually the heat semigroup) which provides an interpolation between the measure in hand and a simpler one, and the analysis amounts to showing inequalities regarding the evolution (e.g., monotonicity of certain quantities). Pathwise-analysis considers a random evolution rather than a deterministic one which allows arguments to be carried out path-by-path rather than on an average sense.

### The logarithmic Laplace transform

The logarithmic Laplace transform is central to our framework: Upper bounds on either itself or its Hessian turn out to be naturally related to functional inequalities of the underlying measure, and it is also used associate between inequalities regarding the center of mass and the covariance structure and relative entropies. In the context of measures on the discrete cube, it was suggested that the log-Laplace transform performs a natural role in concentration inequalities in a paper by Shamir and the second author [ES22]. In the context of log-concave measures on $\mathbb{R}^n$ the log-Laplace transform was long known to perform a central role in concentration inequalities, see [Kla06, KM12, EK11].

### Related techniques in Markov chains

Our technique is also related to Markov chain decomposition techniques ([MR02, JSTV04]) where the analysis of mixing times of a Markov chain is carried out by splitting the chain into

simple-to-analyze components via restriction and projection-like operations. Another common technique in Markov chains is *coupling*, of paths. The coupling technique is also based on path-by-path analysis but otherwise is very different in spirit from our technique, which does not consider paths of the random walk but rather of the measure-evolution process.

#### Localization techniques

The idea of proving concentration inequalities by considering a certain scheme which converts the measures into simpler ones by localizing on space goes back to Gromov and Milman [GM87] and Kannan, Lovász and Simonivits ([KLS95b, KLS95a]). In these works, a measure is effectively decomposed into one-dimensional components called "needles". More recently, needle-decompositions of measures were also used in Riemannian geometry [Kla17, Mil21].

## Summary of notation

- $\rho_{\mathrm{LS}}(\cdot)$ - Modified log-Sobolev (MLSI) coefficient.
- $\mathfrak{gap}(\cdot)$ - Spectral gap.
- $\mathrm{Ent}_\nu[f] := \int f(x)\log f(x) d\nu(x) - \int f d\nu \log\left(\int f d\nu\right)$.
- $\mathrm{D}_{\mathrm{KL}}(\mu||\nu) := \int \log \frac{d\mu}{d\nu}(x) d\mu(x)$.
- For a matrix $A \in \mathbb{M}_{n\times n}$, $\|A\|_{\mathrm{OP}}$ - operator norm; $\rho(A)$ - spectral norm.
- For a measure $\nu$ on $\mathbb{R}^n$:
  - $\mathbf{b}(\nu) := \int x d\nu(x)$ - The **center of mass**.
  - $\mathrm{Cov}(\nu) = \int (x - \mathbf{b}(\nu))^{\otimes 2} d\nu(x)$ - The **covariance matrix**.
  - $\mathbf{Cor}(\nu) := \mathrm{diag}(\mathrm{Cov}(\nu))^{-1/2}\mathrm{Cov}(\nu)\mathrm{diag}(\mathrm{Cov}(\nu))^{-1/2}$, where $\mathrm{diag}(\cdot)$ is the diagonal matrix obtained by setting all the off-diagonal entries to $0$, called the **correlation matrix**.
  - $\mathbf{\Psi}(\nu) := \mathrm{Cov}(\nu)\mathrm{diag}(\mathrm{Cov}(\nu))^{-1}$ the **influence matrix**.
  - $\mathcal{T}_v\nu$ where $v \in \mathbb{R}^n$ is defined as $\frac{d\mathcal{T}_v\nu(x)}{d\nu(x)} := \frac{e^{\langle v,x\rangle} d\nu(x)}{\int e^{\langle v,z\rangle} d\nu(z)}$ (**exponential tilt** of a measure).
  - $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is the standard basis of $\mathbb{R}^n$.
  - $\mathbb{S}^{n-1}$ is the unit sphere in $\mathbb{R}^n$.
- For a measure $\nu$ on $\{-1,1\}^n$,
  - For $u \in \{-1,0,1\}^n$ define $S_u := \left\{x \in \{-1,1\}^n;\ x_i u_i \geq 0,\ \forall i \in [n]\right\}$.
  - $\mathcal{R}_u\nu$ is the normalized restriction of $\nu$ to $S_u$ (**pinning** of $\nu$ according to $u$).
  - For $A \subset [n]$, $\mathcal{R}_A\nu = \mathcal{R}_{\mathbf{1}_A}\nu$ (by abuse of notation, pinning of coordinates in $A$ to $+1$).
  - $P^{\mathrm{GD}}(\nu)$ is the transition kernel for the Glauber-dynamics on $\nu$.

**Acknowledgements**

R.E. would like to thank Kuikui Liu and Bo'az Klartag for enlightening discussions. We would also like to thank Kuikui Liu, Mark Sellke and Eric Vigoda for comments, and Nima Anari and Vishesh Jain for spotting mistakes in an early version of this paper. This work was carried out when R.E. was on a fellowship at the Institute for Advanced Study, Princeton.

# 2 Localization schemes and their associated dynamics

In this section, first we introduce the central object in our framework: a **localization scheme**. Second we discuss the dynamics associated with a localization scheme. Finally, we present several examples of localization schemes.

We fix a set $\Omega$ equipped with a $\sigma$-algebra $\Sigma$. For simplicity the reader may assume that either $\Omega$ is $\mathbb{R}^n$ equipped the $\sigma$-algebra of Lebesgue measurable sets, or $\Omega$ is $\{-1,1\}^n$ equipped with the discrete $\sigma$-algebra. Denote by $\mathcal{M}(\Omega)$ the space of probability measures on $\Omega$.

**Definition 3.** A **localization process** on $\Omega$ is a measure-valued stochastic process $(\nu_t)_{t\geq 0}$ which satisfies the following properties:

(P1) Almost surely, $\nu_t$ is a probability measure on $\Omega$ for all $t$.

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

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

*Remark* 4. In the following, we deal with both **continuous-time** localization processes (namely the time parameter $t$ lives in $[0,\infty)$) and **discrete-time** localization processes (with the discrete time parameter $t = 0, 1, 2, ...$). In order to keep the notation simple, we always assume that $t$ is real-valued, but in the latter case we simply assume that the process is constant in intervals of the form $[k, k+1)$ for $k \in \mathbb{Z}$.

**Definition 5.** A **localization scheme** on $\Omega$ is 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 $\mathcal{L}$ is a localization scheme and $(\nu_t)_{t\geq 0} = \mathcal{L}(\nu)$, then we say that $(\nu_t)_{t\geq 0}$ *is the localization process associated with* $\nu$ *via the localization scheme* $\mathcal{L}$.

## 2.1 Simple examples of localization schemes

To better understand the definition of localization schemes, we give a few simple examples.

### 2.1.1 Example 1: Coordinate-by-coordinate localization

Let $\nu$ be a probability measure on $\{-1,1\}^n$. Define a process $\nu_1, ..., \nu_n$ in the following way: Let $(k_1, ..., k_n)$ be a random variable drawn from the uniform distribution over all permutations of $[n]$. Let $X$ be a random point sampled from $\nu$, independent of $(k_1, ..., k_n)$. For $t \geq 0$, define $\nu_t$ to be the law of $X$ conditioned on $X_{k_1}, ..., X_{k_i}$ for $i = \min(\lfloor t \rfloor, n)$.

Note that, by definition, the sequence $(\nu_t)_{t\geq 0}$ is a martingale, the measures $\nu_t$ are probability measures and almost surely $\nu_t$ is a Dirac measure for all $t \geq n$. Thus, $(\nu_t)_t$ is a localization process. The coordinate-by-coordinate localization scheme is the scheme that assigns the process $(\nu_t)_{t\geq 0}$ to the measure $\nu$.

### 2.1.2 Example 2: Random-subspace localization

Let $\nu$ be a probability measure on $\mathbb{R}^n$ and let $(u_1, ..., u_n)$ be a random orthogonal basis in $\mathbb{R}^n$, distributed according to the Haar measure on the orthogonal group $\mathrm{O}(n)$. Sample $Z$ a random point from $\nu$ independently of $(u_1, ..., u_n)$. For all $t \geq 0$, let $\nu_t$ be the law of $Z$ conditioned on $\langle Z, u_1\rangle, \ldots, \langle Z, u_i\rangle$ for $i = \min(\lfloor t \rfloor, n)$. Then $\nu_t$ is a Doob martingale of probability measures and $\nu_n$ is a Dirac point measure supported on $Z$. Thus, $(\nu_t)_{t\geq 0}$ is a localization process.

### 2.1.3 Example 3: Halfspace-bisection localization

Let $\nu$ be an absolutely continuous probability measure on $\mathbb{R}^n$. Let $\theta_1, \theta_2, \ldots$ be a sequence of independent points generated from the Haar measure on $\mathbb{S}^{n-1}$ and let $\epsilon_1, \epsilon_2, \ldots$ be a sequence of independent $\pm 1$ Bernoulli random variables. Set $\nu_0 = \nu$ and for all $i \in \mathbb{Z}$, define $\nu_i$ inductively as follows. Set

$$t_i = \min\left\{s \in \mathbb{R};\ \ \nu_{i-1}(\{x;\ x \cdot \theta_i \leq s\}) = 1/2\right\},$$

define $H_i = \{x \in \mathbb{R}^n;\ (x \cdot \theta_i - t_i)\epsilon_i \geq 0\}$. For all measurable $A \subset \mathbb{R}^n$, set $\nu_i(A) := 2\nu_{i-1}(A \cap H)$, the normalized restriction of $\nu_{i-1}$ to $H$. In words, $\nu_i$ is generated by choosing a uniformly random direction in the sphere and picking one of the two half-spaces that bisect $\nu_{i-1}$ into two parts of equal mass in that direction. Define $\nu_t = \nu_{\min(\lfloor t \rfloor, n)}$. It is not hard to check that this process is a martingale and thus (P1) and (P2) are satisfied. It can be shown that (P3) is also satisfies, and so the process $(\nu_t)_t$ is a localization process, but we omit the proof since this process is only given for the sake of the example and is not used in the rest of the paper.

### 2.1.4 Example 4: Gaussian channel localization

Consider the case $\Omega = \mathbb{R}^n$. Let $\nu$ be a probability measure on $\Omega$. We construct the process $(\nu_t)_t$ as follows. Let $X \sim \nu$ and let $B_t$ be a standard Brownian motion on $\mathbb{R}^n$ independent of $X$. Define $\nu_t$ to be the law of $X$ conditioned on $tX + B_t$. This process clearly satisfies (P1) and (P2). Under mild conditions we have that $\frac{tX+B_t}{t} = X + \frac{B_t}{t}$ converges to $X$ and that $(\nu_t)_t$ is a localization process. More details about the Gaussian channel localization are provided in section 2.4.

### 2.1.5 Example 5: Subset simplicial-complex localization

Let $k$ be an integer less than or equal to $n$. Let $\Omega$ be the set of all subsets of size of $k$ of $[n]$, and let $\nu$ be a probability measure on $\Omega$. We construct a sequence of measures as follows: Let $X \sim \nu$ and let $j_1, ...j_k$ be a random permutation of $[k]$, independent of $X$. Write $X = \{x_1, .., x_k\}$, where $x_1 < \cdots < x_k$. For $i \in \{1, \ldots, k\}$, define $\nu_i$ to be the law of $X$ conditioned on $X_{j_1}, \ldots, X_{j_i}$. Define $\nu_t = \nu_{\lfloor t \rfloor \wedge k}$.

## 2.2 Dynamics associated with a localization scheme

Given a localization scheme, there is a natural way to introduce a sampling algorithm associated with it. Specifically, starting from a localization process associated with $\nu$ with a localization scheme, we can define a Markov chain whose stationary measure is $\nu$.

**Definition 6.** (**Markov chain associated with a localization process**) Let $(\nu_t)_t$ be a localization process on $\Omega$ and let $\tau > 0$. The dynamics associated with $\nu_t$ and $\tau$ is the Markov chain

whose transition kernel is defined by the formula

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

*Remark* 7. The above definition only makes sense under the condition that $\nu_t$ is almost surely absolutely-continuous with respect to $\nu$, in which case the quantity $\frac{\nu_t(x)}{\nu(x)}$ is well-defined for $\nu$-almost every $x$. Below, we will see how to remedy this issue if this is not the case.

**Fact 8.** *The operator defined in* (1) *is the transition kernel of a reversible Markov chain whose stationary measure is* $\nu$.

*Proof.* By property (P2) of Definition 3, we have $\mathbb{E}[\nu_\tau] = \nu$, which implies that for $\nu$-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 $A, B \subset \Omega$,

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

hence the Markov chain is reversible and has stationary measure $\nu$. □

*Remark* 9. Instead of taking the time $\tau$ to be deterministic, we can take it to be a *stopping time*. It is not hard to verify that we still get a reversible Markov chain.

Another interpretation of the Markov chain associated to localization process is the following: consider a random variable $(X, Y) \in \Omega \times \Omega$ defined by

$$\mathbb{P}(X \in A,\ Y \in B) = \mathbb{E}\left[\nu_\tau(A)\nu_\tau(B)\right]. \tag{2}$$

Define a transition kernel by the formula

$$P_{x\to A} = \mathbb{P}(Y \in A \,|X = x).$$

If $\nu_t$ is absolutely continuous with respect to $\nu$, the conditioning is well-defined for $\nu$-almost every $x$. If that is the case, it is not hard to check that the above transition kernel coincides with the one in definition 6.

Definition 6 introduces a Markov chain with a localization process and a parameter $\tau$. Consequently, every localization scheme gives rise to a sampling algorithm.

**Definition 10.** (**Sampling algorithm associated to a localization scheme**). Given a localization scheme $\mathcal{L}$ on a space $\Omega$ and given $\tau > 0$, we define $P^{(\mathcal{L},\tau)}$ as the mapping from $\mathcal{M}(\Omega)$ to the space of Markov kernels on $\Omega$ where $P^{(\mathcal{L},\tau)}(\nu)$ is given by equation (1).

**Example 1: The coordinate-by-coordinate scheme and Glauber dynamics**

Let us show that the dynamics associated with the coordinate-by-coordinate localization scheme from subsection 2.1.1, with the choice $\tau = n-1$, is the Glauber dynamics, defined as follows.

**Definition 11.** (Glauber dynamics) For a measure $\nu$ on $\{-1,1\}^n$, the Glauber dynamics is the Markov chain whose transition kernel $P^{\mathrm{GD}}(\nu)$ is

$$P^{\mathrm{GD}}_{x\to y}(\nu) = \frac{1}{n}\frac{\nu(y)}{\nu(x)+\nu(y)}\mathbf{1}_{\{\|x-y\|_1=1\}} + p\mathbf{1}_{\{x=y\}},$$

with $p = \frac{1}{n}\sum_{i=1}^n \frac{\nu(x)}{\nu(x)+\nu(x\oplus \mathbf{e}_i)}$ and $x\oplus \mathbf{e}_i$ denotes the vector of $\{-1,1\}^n$ obtaining by flipping the $i$-th coordinate of $x$.

Consider the dynamics obtained from the formula (1) with $\tau = n-1$. First, for $y = x\oplus \mathbf{e}_i$, we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{\nu_\tau(x)\nu_\tau(y)}{\nu(x)}\right]\\
&= \frac{1}{n}\sum_{j\in[n]}\mathbb{E}\left[\frac{\nu_{n-1}(x)\nu_{n-1}(y)}{\nu(x)}\middle| k_n = j\right]\\
&= \frac{\mathbb{P}(\mathrm{supp}(\nu_{n-1})=\{x,y\})}{n}\mathbb{E}\left[\frac{\nu_{n-1}(x)\nu_{n-1}(y)}{\nu(x)}\middle| k_n = i, \mathrm{supp}(\nu_{n-1})=\{x,y\}\right]\\
&= \frac{\nu(x)+\nu(y)}{n\nu(x)}\frac{\nu(x)}{\nu(x)+\nu(y)}\frac{\nu(y)}{\nu(x)+\nu(y)}\\
&= \frac{1}{n}\frac{\nu(y)}{\nu(x)+\nu(y)}.
\end{aligned}$$

Second, if $\|x-y\|_1 > 1$ then $P^{\mathrm{GD}}_{x\to y}(\nu) = 0$, which establishes the fact that the Markov chain associated with the coordinate-by-coordinate process is Glauber dynamics.

*Remark* 12. We can also consider the Markov chain obtained by choosing $\tau = n-\ell$, for some $\ell > 1$. The transition kernel of this Markov chain can be described as follows: Given that the current state is some $x\in\{-1,1\}^n$, the next state will be generated by first choosing a subset of coordinates $A\subset[n]$ with $|A| = \ell$ uniformly at random, and then choosing a point according to the restriction of $\nu$ to the subcube defined by $\{y;\ y_i = x_i,\ \ \forall i\in[n]\setminus A\}$. We refer to this Markov chain as the **$\ell$-Glauber-dynamics**.

**Example 2: The subspace localization scheme and the hit-and-run walk**

Fix a measure $\nu$ on $\mathbb{R}^n$ and let $(\nu_t)_t$ be the subspace-localization process defined in Subsection 2.1.2. Define $\tau = n-1$. Let $(X,Y)$ obey equation (2). Given $x\in\Omega$, consider the expression

$$\begin{aligned}
P_{x\to A} &= \mathbb{P}[Y\in A|X=x]\\
&= \mathbb{E}_{u_1,\ldots,u_{n-1},Z}[\mathbb{P}(Y\in A|X=x,u_1,\ldots,u_{n-1},Z)].
\end{aligned}$$

Note that, by definition, conditioned on $u_1,..,u_{n-1},Z$ we have that $X,Y$ have the law of the restriction of $\nu$ to $\{x;\ \langle x-Z,u_i\rangle = 0,\ \ \forall i\in[n-1]\}$. Therefore, we can understand the

conditioning on $X = x$ as conditioning on the event $E_x = \{\langle Z - x, u_i \rangle = 0, \ \forall i \in [n-1]\}$. In other words

$$\begin{aligned} P_{x \to A} &= \mathbb{E}_{u_1,\dots,u_n} \left[ \mathbb{P} \left( Z \in A | \langle Z - x, u_i \rangle = 0, \ \forall i \in [n-1] \right) \right] \\ &= \mathbb{E}_{u_n} \left[ \mathbb{P} \left( Z \in A | \mathrm{Proj}_{u_n^\perp}(Z - x) = 0 \right) \right]. \end{aligned}$$

In words, given the point $x$, to generate the next point in the Markov chain, we first generate the direction $u_n$ uniformly from the Haar measure on the unit sphere and then generate a point from the restriction of $\nu$ to the fiber $\mathrm{Proj}_{u_n^\perp} x + \mathrm{Span}(u_n)$. This Markov chain is known as the **hit-and-run** walk (see e.g, [Lov99]).

**Other examples of Markov chains**

- In Subsection 2.4.3, we show that the Gaussian-channel localization gives rise to the sampling algorithm introduced in [STL20], where each step is a **restricted Gaussian oracle** step.
- The dynamics associated to the subset simplicial-complex localization in Subsection 2.1.5 with time $\tau = k - \ell$ leads to the $\ell$-**down-up walk** (see [ALOG20]).
- A continuous analogue of the coordinate-by-coordinate localization scheme gives rise to the **coordinate hit-and-run** algorithm (see [NS22]).
- The dynamics associated with the negative-fields localization constructed in Subsection 2.4.4 is the **field dynamics** introduced in [CFYZ21b].

## 2.3 Doob localization schemes

In this subsection, we describe a class of localization schemes which generalizes several examples we have seen so far. Given a set $\Omega$ equipped with a $\sigma$-algebra $\Sigma$, we say that a filtration $(\mathcal{F}_t)_{t \geq 0}$ of $\Omega$ is precise if $\sigma\left(\bigcup_{t \geq 0} \mathcal{F}_t\right) = \Sigma$. Let $\mathbf{F}_\Omega$ be the space of probability measures over the set of precise filtrations on $\Omega$.

Given a distribution $m \in \mathbf{F}_\Omega$, we can introduce a localization scheme using the filtration generated from $m$. For any measure $\nu$ on $\Omega$, we construct a localization process $(\nu_t)_t$ as follows: Let $Y$ be a $\nu$-distributed random variable and let $\mathcal{F}_t$ be a precise filtration generated from $m$ independently of $Y$. $\nu_t$ is defined as follows

$$\nu_t(A) = \mathbb{E}\left[\mathbb{P}(Y \in A | \mathcal{F}_t)\right], \ \ \forall A \subset \Omega \text{ measurable}.$$

It is evident that $\nu_t$ satisfies (P1) and (P2). Since $\mathcal{F}_t$ is precise we also have that $\nu_t$ satisfies (P3). We call localization schemes constructed this way **Doob localization schemes**.

We can verify that coordinate-by-coordinate scheme and the random-subspace localization schemes are Doob localization schemes. For the coordinate-by-coordinate localization scheme, it suffices to take $(\mathcal{F}_t)_{t \geq 0}$ to be $\sigma$-algebra generated by the maps $x \mapsto x_i$ for $i \in k_1, \dots, k_{\lfloor t \rfloor}$. For the random-subspace localization scheme, we take $(\mathcal{F}_t)_{t \geq 0}$ to be $\sigma$-algebra generated by the maps $x \mapsto \langle x, u_i \rangle$ for $i \in k_1, \dots, k_{\lfloor t \rfloor}$.

## 2.4 Linear-tilt-localizations

In this subsection, we describe a family of localization schemes which arise from a process known in the literature under the name **stochastic localization**. To avoid confusion and to make it possible to include new localization schemes to this family, we refer to this family as linear-tilt localizations. The general idea behind these schemes is that, in order to obtain the measure $\nu_{t+dt}$ from the measure $\nu_t$, one multiplies the density by a random linear function. Our definitions in this section make sense whenever $\Omega$ can be naturally embedded in a linear space. However, we will mainly focus our attention on the cases $\Omega = \mathbb{R}^n$ and $\Omega = \{-1,1\}^n \subset \mathbb{R}^n$.

Given a measure $\nu$ on $\Omega$, denote the center of mass of $\nu$ as $\mathbf{b}(\nu) := \int_\Omega x\nu(dx)$.

### 2.4.1 An alternative viewpoint on the coordinate-by-coordinate localization

We begin with an alternative description of the coordinate-by-coordinate localization scheme described in Subsection 2.1.1, which serves as our first example for a linear-tilt localization.

Fix a measure $\nu$ on $\Omega = \{-1,1\}^n$. Let $(k_1, ..., k_n)$ be a uniform permutation of $[n]$ and let $U_1, \ldots, U_n$ be an i.i.d. sequence of independent random variables drawn from the uniform distribution on $[-1,1]$.

Define $\nu_0 = \nu$. For $i = 0, 1, \ldots, n$ define

$$\nu_{i+1}(x) = \nu_i(x)\big(1 + \langle x - \mathbf{b}(\nu_i), Z_i\rangle\big) \tag{3}$$

where $Z_i$ is defined by

$$Z_i = \mathbf{e}_{k_i} \times \begin{cases} \frac{1}{1+\mathbf{b}(\nu_i)_{k_i}} & \text{if } \mathbf{b}(\nu_i)_{k_i} \geq U_i, \\ \frac{-1}{1-\mathbf{b}(\nu_i)_{k_i}} & \text{if } \mathbf{b}(\nu_i)_{k_i} \leq U_i. \end{cases} \tag{4}$$

and where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is the standard basis of $\mathbb{R}^n$ . Finally, to extend the process beyond times $1, \ldots, n$, take $\nu_t = \nu_{\lfloor t \rfloor \wedge n}$.

To see that this is indeed a localization process, first observe that to obtain $\nu_{t+1}$, we multiply $\nu_t$ by a linear function which is equal to 1 at its center of mass, meaning that $\nu_t(\Omega) = \nu_{t+1}(\Omega)$. It is easy to check that $\mathbb{E}[Z_t] = 0$, which verifies (P2). Finally, a direct calculation shows that $\nu_{i+1}$ is in fact a pinning of the $k_i$-th coordinate of $\nu_i$. Indeed, if $x_{k_i} = -\operatorname{sign}(U_i - \mathbf{b}(\nu_i)_{k_i})$ then $\langle x - \mathbf{b}(\nu_i), Z_i\rangle = -1$ in which case the right-hand side of (3) vanishes.

It follows from the above that the definition in this subsection is equivalent to the one given in subsection 2.1.1. This point of view can be used as a motivation for the family of localization schemes considered in this section: Note that it is not evident at all from this definition that the described scheme is a *Doob* localization scheme. However, in this case $(\nu_t)_t$ can be thought of as a Markov chain on the space of measures, where the transition corresponds to multiplication by linear functions with random slopes whose conditional expectation is 0. In this example, the slopes are chosen in the unique way so that it will cause the measure $\nu_{t+1}$ to be a restriction of $\nu_t$ to a subcube.

### 2.4.2 Stochastic localization driven by a Brownian motion

Next we describe the stochastic localization process initially constructed in [Eld13]. Let $\nu$ be a probability measure on $\mathbb{R}^n$, and let $B_t$ be a standard Brownian motion in $\mathbb{R}^n$ adapted to a filtration $\mathcal{F}_t$. Let $(C_t)_t$ be an $\mathcal{F}_t$-measurable process which takes values in the space of $n \times n$

positive-definite matrices. We define a process $(\nu_t)_t$ via the change of measure $\frac{d\nu_t}{d\nu}(x) = F_t(x)$, where the functions $F_t(x)$ solve the following system of stochastic differential equations:

$$F_0(x) = 1, \;\; dF_t(x) = F_t(x)\langle x - \mathbf{b}(\nu_t), C_t dB_t\rangle, \;\; \forall x \in \mathbb{R}^n. \tag{5}$$

The above is an infinite system of differential equations (one for each $x \in \mathbb{R}^n$), but it turns out that a slightly different point of view allows us to view it as a finite system. The uniqueness and existence of the solution of the system of stochastic differential equations are established in [Eld13, EMZ20]. The following fact gives a sufficient condition for the process to be a localization process. It was proved in [Eld13, EMZ20], but we provide a sketch for completeness.

**Fact 13.** *If, for every unit vector $\theta \in \mathbb{S}^{n-1}$, one has $\int_0^\infty |C_t\theta|^2 dt = \infty$ almost surely, then the process $(\nu_t)_t$ is a localization process.*

*Proof.* A direct calculation shows that

$$\begin{aligned} d\nu_t(\mathbb{R}^n) &= \int_{\mathbb{R}^n} dF_t(x)\nu(dx) \\ &= \int_{\mathbb{R}^n} \langle x - \mathbf{b}(\nu_t), C_t dB_t\rangle F_t(x)\nu(dx) \\ &= \left\langle \int x d\nu_t - \mathbf{b}(\nu_t), C_t dB_t \right\rangle = 0. \end{aligned}$$

Moreover, by Ito's formula

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

which shows that $F_t(x)$ is non-negative almost surely for all $x, t$. The above establishes (P1). The martingale property (P2) follows directly from (5). Property (P3) follows from a calculation carried out using Ito's formula which shows that $d\mathrm{Cov}(\nu_t) = -\mathrm{Cov}(\nu_t)C_t^2\mathrm{Cov}(\nu_t)dt +$ martingale , see [Eld20, Section 2]. ∎

A very useful property of the stochastic localization process is the following.

**Fact 14.** *For every $t \geq 0$, we have almost surely that $\nu_t$ attains the form,*

$$\frac{\nu_t(dx)}{\nu(dx)} = \exp\left(Z_t - \frac{1}{2}\langle \Sigma_t x, x\rangle + \langle y_t, x\rangle\right), \tag{6}$$

*where $\Sigma_t = \int_0^t C_s^2 ds$, $y_t = \int_0^t (C_s dB_s + C_s^2 \mathbf{b}(\nu_s)ds)$ and $Z_t$ is a normalizing constant to ensure $\nu_t$ is a probability measure.*

This process turns out to be useful in proving concentration inequalities in both the continuous setting of log-concave measures and the discrete setting, see [Che21, Eld13, Eld20, ES22, LV17, EKZ21].

### 2.4.3 Stochastic localization and sampling via a restricted Gaussian oracle

With the specific choice $C_t = \mathbf{I}_n$, the stochastic localization scheme gives rise, via equation (1), to a natural reversible Markov chain associated with a measure on $\mathbb{R}^n$, which happens to coincide with the restricted Gaussian dynamics given in [STL20, LST21].

For a measure $\nu$ on $\mathbb{R}^n$, $y \in \mathbb{R}^n$ and $\eta > 0$, define the measure $\mathcal{S}_{y,\eta}\nu$ by

$$\mathcal{S}_{y,\eta}\nu(dx) := \frac{\nu(dx)\exp\left(-\frac{1}{2\eta}|x-y|^2\right)}{\int_{\mathbb{R}^n}\exp\left(-\frac{1}{2\eta}|z-y|^2\right)\nu(dz)}$$

**Definition 15.** (The restricted Gaussian dynamics) For a measure $\nu$ on $\mathbb{R}^n$ and $\eta > 0$, the restricted Gaussian dynamics for $\nu$ with parameter $\eta$ is the Markov chain defined by

$$P^{\mathrm{RGD}_\eta}(\nu)_{x\to A} = \mathbb{E}_{y\sim\mathcal{N}(x,\eta\mathbf{I}_n)}\left[\mathcal{S}_{y,\eta}\nu(A)\right].$$

The next proposition is based on an alternative point of view on stochastic localization which was suggested by El Alaoui and Montanari [AM21].

**Proposition 16.** *The dynamics associated to the stochastic localization process with $C_t \equiv \mathbf{I}_n$ and time $\tau$ is the restricted Gaussian dynamics of $\nu$ with $\eta = 1/\tau$.*

*Proof.* It follows from equation (6) that $\nu_t = \mathcal{S}_{y_t,1/t}\nu$, where $y_t = B_t + \int_0^t \mathbf{b}(\nu_s)ds$. Define $X = \lim_{t\to\infty}\frac{1}{t}y_t$. It is shown in [Eld20, Section 2] that $\mathbf{b}(\nu_t)$ almost-surely converges and that $\lim_{t\to\infty}\mathbf{b}(\nu_t)$ has the law $\nu$. Therefore,

$$X = \lim_{t\to\infty}\frac{B_t}{t} + \frac{1}{t}\int_0^t \mathbf{b}(\nu_s)ds = \lim_{t\to\infty}\mathbf{b}(\nu_t) \sim \nu.$$

An application of [AM21, Theorem 2] tells us that $(X, y_t)$ have the same joint law as $(X, tX + \sqrt{t}\Gamma)$ where $\Gamma \sim \mathcal{N}(0, \mathbf{I}_n)$ is independent of $X$. Therefore, if we generate the measure $\nu_t$ according to the process and then generate $X'$ according to $\nu_t$, then the joint law of $(X', \nu_t)$ is the same as that of $(X', \mathcal{S}_{X+\Gamma/\sqrt{t},1/t}\nu)$. In light of equation (2), this completes the proof. □

The last proposition also shows that the stochastic localization process with the choice $C_t = \mathbf{I}_n$ is identical to the Gaussian channel localization from Subsection 2.1.4.

### 2.4.4 The Negative-Fields localization

In this subsection, we introduce a localization scheme which is only relevant for the setting $\Omega = \{-1, 1\}^n$. Roughly speaking, the Negative-Fields localization process associated to a measure $\nu$ on $\{-1, 1\}^n$ is the unique process $(\nu_t)_t$ where $\nu_t$ is almost surely an exponential tilt of the measure by some deterministic external field $v(t)$, and a random pinning of the measure.

Before we move on to the construction of the process, let us introduce some notation. For a measure $\nu$ on $\{-1, 1\}^n$ and for $v \in \mathbb{R}^n$ define

$$\mathcal{T}_v\nu(x) = \frac{e^{\langle x,v\rangle}\nu(x)}{\int e^{\langle y,v\rangle}d\nu(y)},$$

the measure obtained by applying external field $v$ to $\nu$ (or the exponential tilt).

For $u \in \{-1,0,1\}$ we define the $u$-pinning $\mathcal{R}_u\nu$ to be the restriction of $u$ to the subcube

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

By slight abuse of notation, for $A \subset [n]$ define $\mathcal{R}_A\nu(x) = \mathcal{R}_{\mathbf{1}_A}\nu(x)$, namely the measure $\nu$ where the coordinates in $A$ are pinned to $+1$.

We first give a slightly informal account of the construction. Given an initial measure $\nu$ on $\Omega$, the negative-fields localization process is the unique martingale of probability measures, $(\nu_t)_t$, which has the form

$$\nu_t = \mathcal{T}_{-t\vec{\mathbf{1}}}\mathcal{R}_{A_t}\nu,\ \ \forall t > 0,$$

where $\vec{\mathbf{1}} = (1,...,1)$ and $(A_t)_t$ is an almost-surely increasing process of subsets of $[n]$.

A different point of view on the process is to express it as a linear-tilt localization by noting that both the operation of applying an infinitesimal external field and that of pinning amount to linear tilts. Hence, we suppose that for all $t \geq 0$

$$\nu_{t+h}(x) = \nu_t(x)\left(1 + \langle x - \mathbf{b}(\nu_t), Z(t+h) - Z(t)\rangle\right) + o(h), \quad \forall h > 0, \tag{7}$$

where

$$Z_i(t) = -t + \frac{\mathbf{1}\{i \in A_{t+h} \setminus A_t\}}{1 + \mathbf{b}(\nu_t)_i}$$

and the process $(A_t)_t$ of increasing sets is uniquely defined by the equation

$$\mathbb{P}(i \in A_{t+h} \setminus A_t \ \ |A_t) = \mathbf{1}_{\{i \notin A_t\}}\big(1 + \mathbf{b}(\nu_t)_i\big)h + o(h), \quad \forall h > 0.$$

The two last equations render that

$$\mathbb{E}[Z(t+h)|A_t] = Z(t) + o(h), \quad \forall t \geq 0, h > 0$$

which, as we will see, implies that the process is a martingale.

*Remark* 17. Roughly speaking, the measure $\nu_t$ evolves by applying increasingly stronger negative external fields to all sites, and occasionally pinning sites to the value $+1$ in a way that balances the effect of the external fields resulting in a martingale. The dynamics associated with this localization is called the **field dynamics**, and was introduced in [CFYZ21b]. Below, will not study the field dynamics itself, but we will be able to apply this localization to obtain the same results in a way that bypass the dynamics.

The next proposition gives a rigorous account of the above and establishes the existence of a slightly generalized version of the process.

**Proposition 18.** *Let $\nu$ be a probability measure on $\{-1,1\}^n$ and let $v : [0,\infty) \to \mathbb{R}^n$ be a differentiable curve satisfying $v(0) = 0$. There exists a (unique) stochastic process $(u(t))_t$ with $u_t \in \{-1,0,1\}^n$, such that the process $(\nu_t)_t$ defined by $\nu_t := \mathcal{R}_{u(t)}\mathcal{T}_{v(t)}\nu$ is a martingale. This process is uniquely defined by*

$$\mathbb{E}\big(u(t+h)_i - u(t)_i \ \big|u(t)\big) = -\mathbf{1}_{\{u(t)_i=0\}}\big(1 - \operatorname{sign}(v_i'(t))\mathbf{b}(\nu_t)_i\big)v_i'(t)h + o(h),\ \ \forall t \geq 0, i \in [n].$$

*Moreover, we have for all $t \geq 0$ and $x \in \{-1,1\}^n$,*

$$\nu_{t+h}(x) = \nu_t(x)\left(1 + \sum_{i\in[n]}(x_i - \mathbf{b}(\nu_t)_i)\left(hv_i'(t) + \frac{u(t+h)_i - u(t)_i}{1 - \operatorname{sign}(v_i'(t))\mathbf{b}(\nu_s)_i}\right)\right) + o(h). \tag{8}$$

*If, in addition, $\lim_{t\to\infty}|v_i(t)| = \infty$ for all $i \in [n]$, then the above process is a localization process.*

The proof of this proposition is postponed to Appendix A.

### 2.4.5 General linear-tilt localizations

Let us now consider a natural family of localization processes which generalizes the two previous examples. Let $\nu_t$ be a localization process on $\Omega \subset \mathbb{R}^n$ (in general we can consider any linear space). Suppose that for every $t \geq 0$ and $h > 0$ we can write

$$\frac{d\nu_{t+h}}{d\nu}(x) = 1 + \langle x - \mathbf{b}(\nu_t), Z_{t,h} \rangle + o_x(h), \tag{9}$$

such that $\mathbb{E}[Z_{t,h}|\nu_t] = 0$. In other words, the measure $\nu_{t+h}$ is generated from $\nu_t$ via multiplication by a linear function whose slope is a conditionally-centered random variable. Note that the linear functions is equal to 1 at the center of mass of $\nu_t$, which has to be the case given that $\nu_t$ remains a probability measure.

It is not hard to see that the family of localization processes which obey equation (9) contains the coordinate-by-coordinate localization, the stochastic localization and the negative-fields localization.

## 2.5 Spectral gap and modified log-Sobolev inequalities

Consider a localization process $(\nu_t)_t$ obtained from $\nu$ via some localization scheme $\mathcal{L}$, and let $\tau > 0$. Let $P = P^{(\mathcal{L},\tau)}(\nu)$ be the Markov chain obtained by formula (1). The following result establishes functional inequalities regarding the transition kernel $P$ through the analysis of the corresponding localization process.

**Proposition 19.** *If the transition kernel $P$ is the one associated to the localization process $(\nu_t)_t$ via equation (1), then*

$$\mathfrak{gap}(P) = \inf_{\varphi:\Omega\to\mathbb{R}} \frac{\mathbb{E}\left[\mathrm{Var}_{\nu_\tau}[\varphi]\right]}{\mathrm{Var}_\nu[\varphi]}, \text{ and} \tag{10}$$

$$\rho_{\mathrm{LS}}(P) \geq \inf_{f:\Omega\to[0,\infty)} \frac{\mathbb{E}\left[\mathrm{Ent}_{\nu_\tau}[f]\right]}{\mathrm{Ent}_\nu[f]}. \tag{11}$$

*Proof.* For a test function $\varphi : \Omega \to \mathbb{R}$, we have

$$\begin{aligned}
\int_\Omega \varphi(x) \left( \int_\Omega \varphi(y) dP_x(y) \right) d\nu(x) &= \int_\Omega \varphi(x) \left( \mathbb{E}\left[ \frac{\nu_\tau(x) \left( \int_\Omega \varphi(y) d\nu_\tau(y) \right)}{\nu(x)} \right] \right) d\nu(x) \\
&= \mathbb{E}\left[ \int_\Omega \varphi(x) d\nu_\tau(x) \left( \int_\Omega \varphi(y) d\nu_\tau(y) \right) \right] \\
&= \mathbb{E}\left[ \left( \int_\Omega \varphi(x) d\nu_\tau(x) \right)^2 \right].
\end{aligned}$$

Consequently,

$$\begin{aligned}
\frac{\int_{\Omega\times\Omega} (\varphi(x) - \varphi(y))^2 dP_x(y) d\nu(x)}{2\mathrm{Var}_\nu[\varphi]} &= \frac{\int_\Omega \varphi(x)^2 d\nu(x) - \mathbb{E}\left[ \left( \int_\Omega \varphi(x) d\nu_\tau(x) \right)^2 \right]}{\mathrm{Var}_\nu[\varphi]} \\
&= \frac{\mathbb{E}\left[ \int_\Omega \varphi(x)^2 d\nu_\tau(x) - \left( \int_\Omega \varphi(x) d\nu_\tau(x) \right)^2 \right]}{\mathrm{Var}_\nu[\varphi]} \\
&= \frac{\mathbb{E}\left[\mathrm{Var}_{\nu_\tau}[\varphi]\right]}{\mathrm{Var}_\nu[\varphi]}.
\end{aligned}$$

Taking the infimum over all $\varphi$, we obtain (10). For the MLSI, observe that by Jensen's inequality we have for all $x \in \Omega$ and all $f : \Omega \to \mathbb{R}_+$,

$$\mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\int_\Omega f(y)d\nu_\tau(y)\right]\log\mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\int_\Omega f(y)d\nu_\tau(y)\right]$$
$$\leq \mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\left(\int_\Omega f(y)d\nu_\tau(y)\log\left(\int_\Omega f(y)d\nu_\tau(y)\right)\right)\right],$$

where we used the fact that $\mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\right] = 1$. We therefore have

$$\begin{aligned}
&\int_\Omega (Pf)(x)\log((Pf)(x))d\nu(x)\\
&= \int_\Omega \mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\int_\Omega f(y)d\nu_\tau(y)\right]\log\mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\int_\Omega f(y)d\nu_\tau(y)\right]d\nu(x)\\
&\leq \int_\Omega \mathbb{E}\left[\frac{\nu_\tau(x)}{\nu(x)}\left(\int_\Omega f(y)d\nu_\tau(y)\log\left(\int_\Omega f(y)d\nu_\tau(y)\right)\right)\right]d\nu(x)\\
&= \mathbb{E}\left[\left(\int_\Omega \nu_\tau(dx)\right)\int_\Omega f(y)d\nu_\tau(y)\log\left(\int_\Omega f(y)d\nu_\tau(y)\right)\right]\\
&= \mathbb{E}\left[\int_\Omega f(y)d\nu_\tau(y)\log\left(\int_\Omega f(y)d\nu_\tau(y)\right)\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\mathbb{E}\left[\mathrm{Ent}_{\nu_t}[f]\right]}{\mathrm{Ent}_\nu[f]} &= \frac{\mathbb{E}\left[\int_\Omega f(x)\log f(x)d\nu_\tau(x)\right] - \mathbb{E}\left[\int_\Omega f(y)d\nu_\tau(y)\log\left(\int_\Omega f(y)d\nu_\tau(y)\right)\right]}{\mathrm{Ent}_\nu[f]}\\
&\leq \frac{\int_\Omega f(x)\log f(x)d\nu(x)}{\mathrm{Ent}_\nu[f]} - \frac{\int_\Omega (Pf)(x)\log(Pf)(x)d\nu(x)}{\mathrm{Ent}_\nu[f]}\\
&= 1 - \frac{\mathrm{Ent}_\nu[Pf]}{\mathrm{Ent}_\nu[f]}.
\end{aligned}$$

Equation (11) follows by taking infimum over all $f : \Omega \to \mathbb{R}_+$. □

# 3 Approximate conservation of variance and entropy

In this section we describe the main tools used to prove mixing bounds for the dynamics associated with a localization scheme. We introduce the notions of approximate conservation of variance and entropy which, in the context of the coordinate-by-coordinate localization are related in a simple way to spectral independence and entropic independence. In the context of stochastic localization, the concept of variance decay has been applied in order to prove a spectral gap for log-concave measures.

Here, we will see that the same underlying principle is relevant to both techniques, and the difference is in the localization scheme being considered: In both proof techniques, one of the main insights has to do with a connection between the rate of variance decay and the covariance structure of a certain class of measures. Thus our framework captures the relation between covariance structures and variance / entropy decay in both settings using the same derivation.

**Definition 20.** (Approximate Conservation of Variance, discrete time). We say that a localization process $(\nu_i)_i$ satisfies $(\kappa_1, \kappa_2, ...)$-variance conservation up to time $t$, if for every test function $\varphi : \Omega \to \mathbb{R}$ one has

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

The motivation behind this definition is given by the fact that

$$\begin{aligned}\frac{\mathbb{E}[\mathrm{Var}_{\nu_t}[\varphi]]}{\mathrm{Var}_{\nu}[\varphi]} &= \mathbb{E}\left[\prod_{i=1}^{t}\frac{\mathrm{Var}_{\nu_i}[\varphi]}{\mathrm{Var}_{\nu_{i-1}}[\varphi]}\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{t}\mathbb{E}\left[\frac{\mathrm{Var}_{\nu_i}[\varphi]}{\mathrm{Var}_{\nu_{i-1}}[\varphi]}\middle|\nu_{i-1}\right]\right] \geq \prod_{i=1}^{t}(1-\kappa_i).\end{aligned}$$

In light of (10), we immediately obtain the following simple theorem which relates between variance conservation and the spectral gap of the dynamics associated with the localization process.

**Proposition 21.** *If the localization process $(\nu_t)_t$ satisfies $(\kappa_1, \ldots, \kappa_t)$-approximate variance conservation then the dynamics given equation (1) with $\tau = t$ has a spectral gap bounded below by $\prod_{i=1}^{t}(1-\kappa_i)$.*

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

In this section we will present very simple derivation, which gives a powerful tool for proving approximate variance conservation bounds for linear-tilt localization schemes. A special case of this derivation will recover (in a very simple way) the main theorem in the spectral independence framework of [ALOG20].

Suppose that $\Omega \subset \mathbb{R}^n$. Suppose that $(\nu_t)_t$ is a localization process on $\Omega$ whose evolution is given by the equation

$$\nu_{t+1}(x) = \nu_t(x)\left(1 + \langle x - \mathbf{b}(\nu_t), Z_t\rangle\right), \quad \forall x \in \Omega, \tag{12}$$

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

Our first objective is to calculate the variance decay of a test function along this process.

**Claim 22.** *For a test function $\varphi : \Omega \to \mathbb{R}$, we have*

$$\mathbb{E}\left[\mathrm{Var}_{\nu_{t+1}}[\varphi]\middle|\,\nu_t\right] - \mathrm{Var}_{\nu_t}[\varphi] = -\langle v_t, C_t v_t\rangle. \tag{13}$$

*where*

$$v_t := \int_\Omega (x - \mathbf{b}(\nu_t))\varphi(x)\nu_t(dx) \quad \textit{and} \quad C_t := \mathrm{Cov}(Z_t|\nu_t).$$

*Proof.* Fix $\varphi : \Omega \to \mathbb{R}$. We calculate

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

where (i) uses (P2) and (ii) uses the fact that $\mathbb{E}[Z_t|\nu_t] = 0$. By definition of $C_t$ and $v_t$, the right-hand side is equal to $\mathrm{Var}_{\nu_t}[\varphi] - \langle v_t, C_t v_t\rangle$, which completes the proof. ∎

In order to establish an approximate variance conservation bound, we need to give an upper bound on the expression $\langle v_t, C_t v_t\rangle$. We now arrive at what is perhaps the heart of the argument. We use the Cauchy-Schwartz inequality, obtaining

$$\begin{aligned}
\langle v_t, C_t v_t\rangle &= \left|\int_\Omega C_t^{1/2}(x-\mathbf{b}(\nu_t))\varphi(x)\nu_t(dx)\right|^2 \\
&= \sup_{|\theta|=1}\left(\int_\Omega \langle C_t^{1/2}(x-\mathbf{b}(\nu_t)),\theta\rangle\varphi(x)\nu_t(dx)\right)^2 \\
&\le \sup_{|\theta|=1}\int_\Omega \langle C_t^{1/2}(x-\mathbf{b}(\nu_t)),\theta\rangle^2\nu_t(dx)\mathrm{Var}_{\nu_t}[\varphi] \\
&= \|C_t^{1/2}\,\mathrm{Cov}(\nu_t)C_t^{1/2}\|_{\mathrm{OP}}\mathrm{Var}_{\nu_t}[\varphi]. \qquad (14)
\end{aligned}$$

Combining with (13) yields that

$$\frac{\mathbb{E}\left[\mathrm{Var}_{\nu_{t+1}}[\varphi]\middle|\,\nu_t\right]}{\mathrm{Var}_{\nu_t}[\varphi]} \ge 1 - \|C_t^{1/2}\,\mathrm{Cov}(\nu_t)C_t^{1/2}\|_{\mathrm{OP}}. \qquad (15)$$

The derivation above is one of the key insights in the framework presented here: A simple application of the Cauchy Schwartz inequality gives rise to the role of the covariance structure in variance conservation.

Let us now examine how to apply (15) to the case of the coordinate-by-coordinate localization. A simple calculation, using the definition (4) gives that

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

where the last equality is a consequence of the fact that $\int_\Omega x_i^2\nu_t(dx) = 1$. Denoting $D_t$ to be the diagonal matrix whose entries coincide with the diagonal entries of $\mathrm{Cov}(\nu_t)$, we finally have

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

Define for a measure $\tilde{\nu}$ on $\{-1,1\}^n$,

$$\mathbf{Cor}(\tilde{\nu}) := \mathrm{diag}(\mathrm{Cov}(\tilde{\nu}))^{-1/2}\mathrm{Cov}(\tilde{\nu})\mathrm{diag}(\mathrm{Cov}(\tilde{\nu}))^{-1/2}, \tag{16}$$

where $\mathrm{diag}(\cdot)$ is the diagonal matrix obtained by setting all the off-diagonal entries to $0$. So $\mathbf{Cor}(\tilde{\nu})$ is just the matrix of **correlations** between different coordinates of $\tilde{\nu}$. We will also consider the closely-related **influence matrix of the measure**, defined as

$$\mathbf{\Psi}(\nu)_{i,j} := \mathbb{E}_{X\sim\nu}[X_i|X_j = 1] - \mathbb{E}_{X\sim\nu}[X_i|X_j = -1]. \tag{17}$$

The following fact relates the correlation matrix and the influence matrix. Its proof is postponed to Appendix A.

**Fact 23.** *We have* $\mathbf{\Psi}(\nu) = \mathrm{Cov}(\nu)\,\mathrm{diag}(\mathrm{Cov}(\nu))^{-1}$ *and*

$$\|\mathbf{Cor}(\nu)\|_{\mathrm{OP}} = \rho(\mathbf{\Psi}(\nu)) \le \|\mathbf{\Psi}(\nu)\|_{\mathrm{OP}}. \tag{18}$$

In light of the above fact, Equation (15) becomes

$$\frac{\mathbb{E}\left[\mathrm{Var}_{\nu_{t+1}}[\varphi]\middle|\,\nu_t\right]}{\mathrm{Var}_{\nu_t}[\varphi]} \ge 1 - \frac{1}{n-t}\rho(\mathbf{\Psi}(\nu_t)).$$

For $u \in \{-1,0,1\}^n$ define the $u$-pinning of a measure, $\mathcal{R}_u\tilde{\nu}$ to be the restriction of $\tilde{\nu}$ to the sub-cube $S_u$, where

$$S_u := \left\{x \in \{-1,1\}^n;\ x_iu_i \ge 0,\ \forall i \in [n]\right\}. \tag{19}$$

Observe that under the coordinate-by-coordinate localization, $\nu_t$ is of the form $\mathcal{R}_u\nu$ for some $u = u(t)$. The condition that $\|\mathbf{\Psi}(\mathcal{R}_u\nu_t)\|_{\mathrm{OP}}$ is bounded uniformly in $u$ is called **spectral independence**. Plugging the above into (15) and applying Proposition 21 recovers the main theorem in the spectral independence framework [ALOG20]:

**Theorem 24.** *(A reformulation of [ALOG20, Theorem 1.3]) Suppose that $\nu$ is a measure on $\{-1,1\}^n$ such that for all $u \in \{-1,0,1\}^n$,*

$$\rho(\mathbf{\Psi}(\mathcal{R}_u\nu)) \le \eta_{|u|_1}.$$

*Then the spectral gap of the $k$-Glauber dynamics on $\nu$ is at least $\prod_{i=0}^{n-k-1}(1-\frac{\eta_i}{n-i})$.*

*Remark* 25. There are some delicate differences between the above theorem and [ALOG20, Theorem 1.3]. In the latter, $\tilde{\eta}_i$ is taken to be the operator norm of the matrix $\mathbf{\Psi}(\mathcal{R}_i\nu) - \mathbf{I}_n$ rather than that of the matrix $\mathbf{\Psi}(\nu)$. Precisely, their result states that spectral gap is bounded from below by the expression $\frac{1}{n}\prod_{i=0}^{n-k-1}(1-\frac{\tilde{\eta}_i}{n-i-1})$ where $\tilde{\eta}_i = \eta_i - 1$. So the extra factor $1/n$ is traded for the fact that $\eta_i$ can be replaced by $\tilde{\eta}_i = \eta_i - 1$.

In our proof, we completely bypassed the need to use the notion of high-dimensional expanders or the up-down walk. The inequalities relating different levels of the up-down walk were replaced by a simple application of the Cauchy-Schwartz inequality.

### 3.1.1 The case of stochastic localization; bounds for the spectral gap of log-concave measures

The same derivation (almost verbatim) as above gives a similar bound for the variance decay of the stochastic localization process (5). Here we briefly outline the argument which can be found in [Eld].

Let $\nu_t$ be the stochastic localization process associated to a measure $\nu$ on $\mathbb{R}^n$, via equation (5) with some driving matrix $C_t$. Fix a test function $\varphi : \mathbb{R}^n \to \mathbb{R}$. The continuous analogue of (13) takes the form

$$d\mathrm{Var}_{\nu_t}[\varphi] = -\left|\int C_t^{1/2}(x - \mathbf{b}(\nu_t))\varphi(x)d\nu_t(x)\right| dt + \text{martingale}.$$

Equation (14) (or, in other words, an application of Cauchy Schwartz) gives

$$\left|\int C_t^{1/2}(x - \mathbf{b}(\nu_t))\varphi(x)d\nu_t(x)\right| \leq \|C_t^{1/2}\mathrm{Cov}(\nu_t)C_t^{1/2}\|_{\mathrm{OP}}\mathrm{Var}_{\nu_t}[\varphi]$$

which implies that

$$d\mathrm{Var}_{\nu_t}[\varphi] \geq -\|C_t^{1/2}\mathrm{Cov}(\nu_t)C_t^{1/2}\|_{\mathrm{OP}}\mathrm{Var}_{\nu_t}[\varphi]dt + \text{martingale}$$

By integration, we get the variance decay bound

$$\frac{\mathbb{E}[\mathrm{Var}_{\nu_t}[\varphi]]}{\mathrm{Var}_{\nu}[\varphi]} \geq \mathbb{E}\left[e^{\int_0^t \|C_s^{1/2}\mathrm{Cov}(\nu_s)C_s^{1/2}\|_{\mathrm{OP}}ds}\right]^{-1} \tag{20}$$

Therefore, upper bounds on the process $t \to \|\mathrm{Cov}(\nu_t)\|_{\mathrm{OP}}$ correspond to approximate conservation of variance bounds.

A classical inequality by Brascamp and Lieb allows us to relate the Dirichlet form $\mathbb{E}_\nu[|\nabla\varphi|^2]$ to $\mathbb{E}[\mathrm{Var}_{\nu_t}[\varphi]]$. Namely, due to (6) we have that the measure $\nu_t$ is $\alpha_t$-strongly log-concave with $\alpha_t$ being the smallest eigenvalue of $\Sigma_t = \int_0^t C_s^2 ds$, which implies that

$$\mathbb{E}[\mathrm{Var}_{\nu_t}[\varphi]] \leq \frac{1}{\alpha_t}\mathbb{E}_\nu[|\nabla\varphi|^2].$$

Combining this with (20) gives a way to obtain lower bounds for the spectral gap of a measure $\nu$ on $\mathbb{R}^n$ via the analysis of the process $t \to \|\mathrm{Cov}(\nu_t)\|_{\mathrm{OP}}$. This is one of the main ideas underlying the stochastic localization technique, which in particular led to the near-solution, due to the first author ([Che21]), of the Kannan-Lovász-Simonovits conjecture ([KLS95a]) and Bourgain's slicing problem.

## 3.2 Approximate conservation of entropy

It turns out that there are also natural sufficient conditions regarding the conservation of the entropy along a localization process. These conditions are closely related to the notion of **entropic independence** put forth in [AJK+21b].

**Definition 26.** (Entropy conservation, discrete time). We say that a localization process $(\nu_i)_i$ satisfies $(\kappa_1, \kappa_2, ...)$-entropy conservation up to time $t$, if for every test function $f : \Omega \to \mathbb{R}_+$ one has

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

A telescopic-product argument, completely analogous to the one in the proof of Proposition 21, together with equation (11), yields the following.

**Proposition 27.** *If the localization process $(\nu_t)_t$ satisfies a $(\kappa_1, \ldots, \kappa_t)$-entropy conservation then the dynamics given equation (1) with $\tau = t$ has a MLSI coefficient bounded below by $\prod_{i=1}^{t}(1-\kappa_i)$.*

*Remark* 28. We refer the reader to [AJK+21b, Section 1.1] for some examples where this bound can be applied. Usually, however, the expression $\prod_{i=1}^{t}(1-\kappa_i)$ will give optimal bounds. Below we will see better ways to extracting the power of approximate conservation of entropy.

### 3.2.1 Entropic stability

Next, we discuss the case $\Omega \subset \mathbb{R}^n$ (relevant in particular to $\Omega = \mathbb{R}^n$ and $\Omega = \{-1, 1\}^n$). We introduce a notion which turns out to be very useful in controlling the decay of entropy along linear-tilt localizations.

For a measure $\nu$ on $\mathbb{R}^n$ and $v \in \mathbb{R}^n$, define the exponential tilt $\mathcal{T}_v\nu$ by

$$\frac{d\mathcal{T}_v\nu(x)}{d\nu(x)} := \frac{e^{\langle v,x\rangle}}{\int e^{\langle v,z\rangle} d\nu(z)}.$$

A central definition in our framework is the following.

**Definition 29.** (Entropic stability). For $\nu$ on $\Omega \subset \mathbb{R}^n$, a function $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ and $\alpha > 0$, we say that $\nu$ is $\alpha$-entropically stable with respect to $\psi$ if

$$\psi(\mathbf{b}(\mathcal{T}_v\nu), \mathbf{b}(\nu)) \leq \alpha \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_v\nu||\nu), \quad \forall v \in \mathbb{R}^n.$$

In order for the definition to make sense, $\psi(x, y)$ needs to vanish when $x = y$. For the sake of intuition, we may think of the case $\psi(x, y) = |x - y|^2$. Roughly speaking, entropic stability amounts to the fact that the center of mass does not move much when tilting the measure in terms of the relative entropy of the tilt.

What follows is a very useful observation, based the **principle of maximum entropy**. This observation is essentially due to [AJK+21b], where it was used in the context of entropic independence, which can be understood as a specific case of entropic stability.

**Fact 30.** *For every measure $\nu$ on $\Omega$ and every function $g : \mathbb{R}^n \to \mathbb{R}$, one has*

$$\inf_{\mu} \frac{\mathrm{D}_{\mathrm{KL}}(\mu||\nu)}{g(\mathbf{b}(\mu))} = \inf_{v \in \mathbb{R}^n} \frac{\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_v\nu||\nu)}{g(\mathbf{b}(\mathcal{T}_v\nu))}. \tag{21}$$

*Proof.* By the maximum entropy principle, the minimum over measures $\mu$ of the relative entropy $\mathrm{D}_{\mathrm{KL}}(\mu||\nu)$ under the linear constraint $\mathbf{b}(\mu) = b_0$ is attained for a measure such that $\frac{d\mu}{d\nu} \propto \exp(L(\cdot))$ for some linear function $L : \mathbb{R}^n \to \mathbb{R}$, or in other words the optimal $\mu$ is of the form $\mathcal{T}_v\nu$ for some $v \in \mathbb{R}^n$. ∎

An immediate corollary is the following,

**Lemma 31.** *Suppose that a measure $\nu$ is $\alpha$-entropically stable with respect to $\psi$. Then, for every measure $\mu$ which is absolutely continuous with respect to $\nu$, we have*

$$\psi(\mathbf{b}(\mu), \mathbf{b}(\nu)) \leq \alpha \mathrm{D}_{\mathrm{KL}}(\mu||\nu).$$

*Remark* 32. In view of the last lemma, we see that the notion of entropic stability is closely related to **transportation-entropy** bounds introduced by Talagrand (see [GL10] for a survey). While entropic stability refers to the fact that the center of mass does not move under change of measure with small relative entropy, a transportation-entropy inequality alludes to the transportation distance being small under such change of measure. Thus, in a sense entropic stability bounds are a weaker version of transportation-entropy bounds.

One last ingredient which will be useful to us is the following formula which can be obtained via the logarithmic Laplace transform.

**Lemma 33.** *([BE19, Lemma 1]). Let $\nu$ be a measure on $\mathbb{R}^n$ such that $\mathrm{Cov}(\nu)$ is invertible. Denote by $K \subset \mathbb{R}^n$ the interior of the convex hull of the support of $\nu$. Then there exists a unique function $v : K \to \mathbb{R}^n$ such that*

$$\mathbf{b}(\mathcal{T}_{v(x)}\nu) = x, \quad \forall x \in K.$$

*Moreover, denoting $g(x) := \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{v(x)}\nu||\nu)$, we have*

$$\nabla g(\mathbf{b}(\nu)) = 0, \quad \nabla^2 g(x) = \mathrm{Cov}(\mathcal{T}_{v(x)}\nu)^{-1}, \quad \forall x \in K. \tag{22}$$

*Finally, if $L_\nu(\theta) := \log \int e^{\langle x,\theta\rangle} d\nu(x)$ is the Logarithmic Laplace transform of $\nu$, then we have that*

$$g(x) = \sup_{\theta \in \mathbb{R}^n} \langle x, \theta\rangle - L_\nu(\theta)$$

*hence $g$ is the Legendre dual of $L_\nu$.*

*Remark* 34. The role of the logarithmic Laplace transform in concentration inequalities on the discrete hypercube was suggested in [ES22]. The boundedness of its hessian is closely related to the notion of fractional log-concavity of the characteristic polynomial, introduced in [AASV21].

### 3.2.2 The case of the coordinate-by-coordinate localization: Entropic independence

As a warm-up, in this section we essentially recover the results of [AJK+21b], deriving a natural criterion for approximate conservation of entropy for $\Omega = \{-1, 1\}^n$ with the coordinate-by-coordinate localization process. The proof here boils down to ideas which are quite similar to the ones that appear in [AJK+21b], with two crucial differences which will (arguably) make our argument cleaner: 1. We directly analyze measures on the hypercube rather than measures on $\binom{n}{k}$. 2. The notion of *fractional log-concavity* is replaced by a definition which involves the logarithmic Laplace transform of the measure, which seems to arise naturally in this context, and due to which the proof involves much simpler formulas.

For $x, y \in [-1, 1]^n$, define

$$H(x, y) = \sum_{i=1}^{n} \mathbf{1}_{|y_i|<1} \left( \frac{1 + x_i}{2} \log\left(\frac{1 + x_i}{1 + y_i}\right) + \frac{1 - x_i}{2} \log\left(\frac{1 - x_i}{1 - y_i}\right)\right), \tag{23}$$

It turns out that in the context of the coordinate-by-coordinate scheme, the key will be to consider $\alpha$-entropic stability with respect to the function $\psi(x, y) = H(x, y)$. With this choice, entropic stability is essentially the same is **entropic independence** introduced in [AJK+21b], and the following proposition shows that it implies approximate conservation of entropy under the coordinate-by-coordinate localization.

**Proposition 35.** *Let $\nu_t$ be the coordinate-by-coordinate localization process for some measure $\nu$ on $\{-1,1\}^n$. For every $t$, if the measure $\nu_t$ is $\kappa$-entropically stable with respect to the function $\psi(x,y) = H(x,y)$ then we have the approximate conservation of entropy bound*

$$\mathbb{E}\left[\mathrm{Ent}_{\nu_{t+1}}[f]\,|\nu_t\right] \geq \left(1 - \frac{\kappa}{n-t}\right)\mathrm{Ent}_{\nu_t}[f],$$

*for all $f : \Omega \to \mathbb{R}_+$ and $t \in [n-1]$. Consequently, if $\mathcal{R}_u\nu$ is $\kappa$-entropically stable with respect to the same function, for all $u \in \{-1,0,1\}^n$, then the $\ell$-Glauber dynamics $P = P^{\mathrm{GD},\ell}(\nu)$ has MLSI coefficient*

$$\rho_{\mathrm{LS}}(P) \geq \prod_{i=0}^{n-\ell-1}\left(1 - \frac{\kappa}{n-i}\right).$$

*Proof.* Fix a probability measure $\nu$ on $\Omega$ and a function $f : \Omega \to \mathbb{R}_+$. Let $(\nu_t)_t$ be the coordinate-by-coordinate localization process of $\nu$. The derivation obtained in Subsection 2.4.1 gives the following: Conditioned on $\nu_t$, we have

$$\begin{aligned}\int f(x)d\nu_{t+1}(x) &\overset{(3)}{=} \int f(x)\left(1 + \langle x - \mathbf{b}(\nu_t), Z\rangle\right)d\nu_t(x)\\ &= \int f(x)d\nu_t(x)\left(1 + \left\langle\left(\frac{\int xf(x)d\nu_t(x)}{\int f(x)\nu_t(dx)} - \mathbf{b}(\nu_t)\right), Z\right\rangle\right)\\ &= \nu_t(f)\left(1 + \langle V, Z\rangle\right) \qquad (24)\end{aligned}$$

where

$$V := \frac{\int xf(x)\nu_t(dx)}{\int f(x)\nu_t(dx)} - \mathbf{b}(\nu_t)$$

and

$$Z = \mathbf{e}_{k_{t+1}} \times \begin{cases}\frac{1}{1+\mathbf{b}(\nu_t)_{k_{t+1}}} & \text{with probability } \frac{1+\mathbf{b}(\nu_t)_{k_{t+1}}}{2},\\ \frac{-1}{1-\mathbf{b}(\nu_t)_{k_{t+1}}} & \text{with probability } \frac{1-\mathbf{b}(\nu_t)_{k_{t+1}}}{2}.\end{cases}$$

We have

$$\mathbb{E}\left[(1+\langle V,Z\rangle)\log(1+\langle V,Z\rangle)|\nu_t\right] = \frac{1}{n-t}\sum_{i\in[n]\setminus\{k_1,\ldots,k_t\}} q(t;i)$$

where

$$q(t;i) = \mathbb{E}\left[(1+\langle V,Z\rangle)\log(1+\langle V,Z\rangle)|\nu_t, k_{t+1} = i\right].$$

Now, writing $\frac{d\mu_t}{d\nu_t}(x) = \frac{f(x)}{\int f(z)d\nu_t(z)}$, we have $V = \mathbf{b}(\mu_t) - \mathbf{b}(\nu_t)$ and

$$\begin{aligned}q(t;i) &= \frac{1+b(\nu_t)_i}{2}\left(1 + \frac{V_i}{1+\mathbf{b}(\nu_t)_i}\right)\log\left(1 + \frac{V_i}{1+\mathbf{b}(\nu_t)_i}\right)\\ &\quad + \frac{1-b(\nu_t)_i}{2}\left(1 - \frac{V_i}{1-\mathbf{b}(\nu_t)_i}\right)\log\left(1 - \frac{V_i}{1-\mathbf{b}(\nu_t)_i}\right)\\ &= \frac{1+\mathbf{b}(\mu_t)_i}{2}\log\left(\frac{1+\mathbf{b}(\mu_t)_i}{1+\mathbf{b}(\nu_t)_i}\right) + \frac{1-\mathbf{b}(\mu_t)_i}{2}\log\left(\frac{1-\mathbf{b}(\mu_t)_i}{1-\mathbf{b}(\nu_t)_i}\right).\end{aligned}$$

Combining the last three displays, we have

$$\mathbb{E}\left[(1+\langle V,Z\rangle)\log(1+\langle V,Z\rangle)|\nu_t\right] = \frac{1}{n-t}H(\mathbf{b}(\mu_t),\mathbf{b}(\nu_t)). \qquad (25)$$

We finally get

$$\mathbb{E}[\mathrm{Ent}_{\nu_{t+1}}[f]\big|\ \nu_t] = \mathrm{Ent}_{\nu_t}[f] - \frac{1}{n-t}\nu_t(f)H(\mathbf{b}(\mu_t),\mathbf{b}(\nu_t)).$$

The assumption that $\nu_t$ is $\kappa$-entropically stable with respect to $h$, combined with Lemma 31, amounts to

$$H(\mathbf{b}(\mu_t),\mathbf{b}(\nu_t)) \le \kappa \mathrm{D}_{\mathrm{KL}}(\mu||\nu) = \frac{\kappa}{\nu_t(f)}\mathrm{Ent}_{\nu_t}[f].$$

Combining the last two displays gives

$$\mathbb{E}\left[\mathrm{Ent}_{\nu_{t+1}}[f]\big|\,\nu_t\right] \ge \mathrm{Ent}_{\nu_t}[f] - \frac{\kappa}{n-t}\mathrm{Ent}_{\nu_t}[f],$$

By definition of approximate entropy conservation, this completes the proof of the first part. The second part follows immediately by the use of Proposition 27. □

In the paper [AJK+21b], the authors show that entropic independence is naturally related to the log-concavity of a power of the characteristic polynomial, referred to as *fractional log-concavity*. Next, we show that a somewhat analogous, very simple derivation, gives that entropic stability with respect to the function $h$ is implied by the fact that exponential tilts of the measure have a bounded correlation matrix. The proof of the next lemma is found in Appendix A

**Lemma 36.** *Suppose that a measure $\nu$ on $\{-1,1\}^n$ satisfies*

$$\|\mathbf{Cor}(\mathcal{T}_v\nu)\|_{\mathrm{OP}} \le \alpha,\ \ \forall v \in \mathbb{R}^n \tag{26}$$

*where $\mathbf{Cor}(\cdot)$ is defined in (16). In this case, $\nu$ is $\alpha$-entropically stable with respect to $\psi(x,y) = H(x,y)$.*

*Remark* 37. Fact 23 shows that $\alpha$-spectral independence is equivalent to the fact that $\mathrm{Cov}(\nu) \preceq (\alpha+1)\,\mathrm{diag}(\mathrm{Cov}(\nu))$. Therefore, the condition given in the lemma can be thought of as spectral independence for all tilts.

*Remark* 38. The condition (26) is the exact condition given in [ES22], where it is shown that it implies nontrivial concentration of Lipchitz functions.

### 3.2.3 Entropic decay for stochastic localization

Let $\nu$ be a measure on $\mathbb{R}^n$ and consider the stochastic localization process $(\nu_t)_t$ defined by (5). Fix a non-zero, measurable function $f : \Omega \to \mathbb{R}_+$. For every $t > 0$, define a probability measure $\mu_t$ by the equation

$$\frac{d\mu_t}{d\nu_t}(x) = \frac{f(x)}{\int f(x)\nu_t(dx)}.$$

Consider the martingale

$$M_t := \nu_t(f) = \int f(x)F_t(x)\nu(dx).$$

Using equation (5), we can calculate

$$\begin{aligned} dM_t &= \int f(x)F_t(x)\langle x - \mathbf{b}(\nu_t), C_t dB_t\rangle \nu(dx) \\ &= M_t\big\langle C_t(\mathbf{b}(\mu_t) - \mathbf{b}(\nu_t)), dB_t\big\rangle. \end{aligned}$$

Using Itô's formula, we have

$$\begin{aligned} d(M_t \log M_t) &= \frac{d[M]_t}{2M_t} + \text{ martingale} \\ &= \frac{1}{2} M_t |C_t(\mathbf{b}(\mu_t) - \mathbf{b}(\nu_t))|^2 dt + \text{ martingale}. \end{aligned}$$

We finally obtain

$$\begin{aligned} d\mathrm{Ent}_{\nu_t}[f] &= d \int f(x) \log f(x) \nu_t(dx) - d(M_t \log M_t) \\ &= -\frac{1}{2} \nu_t(f) \left| C_t \left( \mathbf{b}(\mu_t) - \mathbf{b}(\nu_t) \right) \right|^2 dt + \text{martingale}. \end{aligned} \tag{27}$$

We arrive at the following,

**Proposition 39.** *For a fixed $T > 0$, suppose that, almost surely for all $t \in [0, T]$ that $\nu_t$ is $\alpha_t$-entropically stable with respect to the function $\psi(x, y) = \frac{1}{2}|C_t(x - y)|^2$. Then we have the approximate entropic conservation bound*

$$\mathbb{E}\left[\mathrm{Ent}_{\nu_T}[f]\right] \geq e^{-\int_0^T \alpha_t dt} \mathrm{Ent}_{\nu}[f]. \tag{28}$$

*Proof.* The entropic stability assumption combined with Lemma 31 and equation (27) yields that

$$d\mathrm{Ent}_{\nu_t}[f] \geq -\alpha_t \mathrm{Ent}_{\nu_t}[f] dt + \text{martingale}, \quad \forall t \in [0, T].$$

Consequently, by applying Ito's formula, we have that the process $t \to e^{\int_0^t \alpha_s ds} \mathrm{Ent}_{\nu_t}[f]$ is a submartingale. Therefore,

$$\mathbb{E}\left[ e^{\int_0^T \alpha_t dt} \mathrm{Ent}_{\nu_T}[f] \right] \geq \mathrm{Ent}_{\nu}[f],$$

which completes the proof. □

The following lemma is useful for establishing entropic stability with respect to quadratic functions.

**Lemma 40.** *Let $\nu$ be a measure on $\mathbb{R}^n$ and $C, A$ be positive-definite matrices. Suppose that for every $v \in \mathbb{R}^n$ one has*

$$\mathrm{Cov}(\mathcal{T}_v \nu) \preceq A, \tag{29}$$

*Then $\nu$ is $\alpha$-entropically stable with respect to the function $\psi(x, y) = \frac{1}{2}|C(x - y)|^2$, for $\alpha = \|CAC\|_{\mathrm{OP}}$.*

*Proof.* Define $v(x)$, $g(x) := \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{v(x)}\nu || \nu)$ and $K$ as in Lemma 33. We have,

$$\nabla^2 g(x) \overset{(22)}{=} \mathrm{Cov}(\mathcal{T}_{v(x)}\nu)^{-1} \overset{(29)}{\succeq} A^{-1}, \quad \forall x \in K,$$

and $\nabla g(\mathbf{b}(\nu)) = 0$. Define $h(x) = \frac{1}{2}|C(x - \mathbf{b}(\nu))|^2$. Then $\nabla h(\mathbf{b}(\nu)) = 0$ and, for all $x$, $\nabla^2 h(x) = C^2$. Therefore, for all $x \in K$, we have

$$\|CAC\|_{\mathrm{OP}} \nabla^2 g(x) \succeq \|CAC\|_{\mathrm{OP}} A^{-1} \succeq C^2 = \nabla^2 h(x).$$

Since $g(\mathbf{b}(\nu)) = h(\mathbf{b}(\nu)) = 0$ and $\nabla g(\mathbf{b}(\nu)) = \nabla h(\mathbf{b}(\nu)) = 0$, $g$ and $h$ coincide up to first order Taylor expansion around $\mathbf{b}(\nu)$. It follows that $\|CAC\|_{\mathrm{OP}} g(x) \geq h(x)$. Together with Fact 30, we complete the proof. □

### 3.2.4 Entropic decay for the negative-fields localization

It turns out that the entropy decay for the negative-fields localization is also governed by the function $H(x,y)$ in equation (23).

**Proposition 41.** *Consider a localization process $\nu_t$ obtained via the negative-fields localization scheme (with the choice $v(t) = -t\vec{\mathbf{1}}$). Let $f : \{-1,1\}^n \to \mathbb{R}$ and let $\mu_t$ be the measure obtained via the formula $\frac{d\mu_t(x)}{d\nu_t(x)} = f(x)$. Suppose that $\nu_t$ is $\alpha$-entropically stable with respect to the function $H(x,y)$ defined in equation (23), then we have the approximate entropy conservation bound*

$$\mathbb{E}[\mathrm{Ent}_{\nu_{t+h}}[f]|\nu_t] \geq \mathrm{Ent}_{\nu_t}[f](1 - 4h\alpha) + o(h), \quad \forall t, h \geq 0.$$

The proof of the above proposition is just a calculation, in the same spirit as the derivation of entropic decay for the coordinate-by-coordinate process and stochastic localization. The proof is found in Appendix A.

## 3.3 Entropic stability via spectral independence

Lemma 36 and Lemma 40 both rely on the logarithmic Laplace transform to show that entropic-stability can be deduced from bounds on the covariance matrix of different tilts of the measure. On the other hand, the weaker notion of spectral independence only requires a corresponding bound on the covariance matrix itself, and in order to get a spectral gap via the spectral independence framework, we need a bound on the influence matrix for all pinnings rather than all tilts (see remark 37).

It turns out that requiring all pinnings of a measure $\nu$ to be spectrally-independent is sufficient for entropic stability with respect to a quadratic function. Recall the definition of $\mathbf{\Psi}(\nu)$ from equation (16) and recall that for all $u \in \{-1,0,1\}^n$, $\mathcal{R}_u\nu$ is defined to be the restriction of $\nu$ to the subcube $S_u$ (defined in (19)).

**Theorem 42.** *Let $\nu$ be a probability measure on $\{-1,1\}^n$ and let $\alpha \geq 1$. Suppose that*

$$\|\mathbf{Cor}(\mathcal{R}_u\nu)\|_{\mathrm{OP}} \leq \alpha, \quad \forall u \in \{-1,0,1\}^n. \tag{30}$$

*Then $\nu$ is $8\alpha$-entropically stable with respect to $\psi(x,y) = \frac{1}{2}|x-y|^2$.*
*Furthermore, if for some constants $K \geq 1, C \geq 1$, for every $i \in [n]$ and for every $u, w \in \{-1,0,1\}^n$ with $\mathrm{supp}(u) \cap \mathrm{supp}(w) = \emptyset$ with $u_i = w_i = 0$, we have*

$$\frac{1 + \mathbf{b}_i(\mathcal{R}_w\mathcal{R}_u\nu)}{1 - \mathbf{b}_i(\mathcal{R}_w\mathcal{R}_u\nu)} \leq K\frac{1 + \mathbf{b}_i(\mathcal{R}_u\nu)}{1 - \mathbf{b}_i(\mathcal{R}_u\nu)}, \text{ and} \tag{31}$$

$$1 - \mathbf{b}_i(\mathcal{R}_u\nu) \geq \frac{1}{C}, \tag{32}$$

*then $\nu$ is $768\alpha K^3 C$-entropically stable with respect to the function $H(x,y)$ as defined in equation (23).*

The proof is found in Section 6.

Recall the $\ell$-Glauber dynamics (see remark 12 above), which is associated to the coordinate-by-coordinate localization scheme with stopping time $\tau = n - \ell$. Denote its transition kernel by $P^{\ell-\mathrm{GD}}(\nu)$.

The above theorem will allow us to recover the extension of the spectral independence framework due to Chen, Liu and Vigoda ([CLV21]), and show that spectral independence for

all pinnings implies an MLSI for the $\ell$-Glauber dynamics with $\ell = \Omega(n)$. This is true under the extra condition that the marginals of the measure are bounded away from $-1$ and $1$ under pinnings:

**Definition 43.** (bounded marginals). We say that a measure $\nu$ on $\mathbb{R}^n$ has $b$-marginally bounded if for all $i \in [n]$ and all $u \in \{-1, 0, 1\}^n$ with $u_i = 0$, we have

$$|\mathbf{b}(\mathcal{R}_u \nu)_i| \leq 1 - b.$$

The following theorem essentially recovers the framework of [CLV21].

**Theorem 44.** *Let $\nu$ be a probability measure on $\{-1, 1\}^n$ and let $\alpha \geq 1$. Suppose that*

$$\rho(\mathbf{\Psi}(\mathcal{R}_u \nu)) \leq \alpha, \quad \forall u \in \{-1, 0, 1\}^n. \tag{33}$$

*Moreover, suppose that $\nu$ is $b$-marginally bounded. Then for every $\varepsilon \in \left(\frac{8a}{bn}, \frac{1}{2}\right)$,*

$$\rho_{\mathrm{LS}}(P^{\ell-\mathrm{GD}})(\nu) \geq \varepsilon^{C\alpha/b}.$$

*where $\ell = \lceil \varepsilon n \rceil$ and $C > 0$ is a universal constant.*

*Remark* 45. This result also holds under a weaker condition than having bounded-marginals, namely it is enough that the measure has $(1/b)$-tame marginals, see definition 59 below.

In [CLV21], a clever argument is used to show that when the model has bounded degrees (see [CLV21] for a precise definition) and $\varepsilon$ is a small enough constant (as a function of the maximum degree), then mixing of the $\ell$-Glauber dynamics for $\ell = \varepsilon n$ implies mixing of the 1-Glauber dynamics. In fact, it is even easier to show that in the case of bounded degrees, one may sample directly by using the $\ell$-Glauber dynamics, as the single step has complexity bounded by $n^{o(1)}$.

*Proof of Theorem 44.* The bounded marginals assumption implies that the condition (31) holds true for any pinning of the measure $\nu$ with $K = 2/b$. Using Theorem 42, we have that for every $u \in \{-1, 0, 1\}^n$, $\mathcal{R}_u \nu$ is $(48\alpha/b)$-entropically stable with respect to $H(x, y)$ defined in (23). An application of Proposition 35 gives that the $\ell$-Glauber dynamics has MLSI

$$\begin{aligned}
\rho_{\mathrm{LS}}(P^{\ell-\mathrm{GD}})(\nu) &\geq \prod_{i=0}^{n-\ell-1} \left(1 - \frac{48\alpha b^{-1}}{n-i}\right) \\
&= \exp\left(\sum_{i=0}^{\lceil (1-\varepsilon)n \rceil} \log\left(1 - \frac{100\alpha b^{-1}}{n-i}\right)\right) \\
&\overset{(i)}{\geq} \exp\left(-\sum_{i=0}^{\lceil (1-\varepsilon)n \rceil} \frac{200\alpha b^{-1}}{n-i}\right) \geq \varepsilon^{Ca/b},
\end{aligned}$$

where (i) uses the assumption that $\varepsilon > \frac{8a}{bn}$ and the fact that $\log(1 - x) \geq -2x$ for $x > -1/2$. This completes the proof. ∎

# 4 Annealing via a localization scheme

A powerful tool in our framework will be to conceptually concatenate two localization schemes, running one localization scheme up to some time $t$ and then running a second scheme where $\nu_t$ is used as the initial condition. As we will see, in many cases this tool will allow us to use the first stage as an annealing procedure which tames the measure in a way that provides a good starting point for the second stage.

Consider two localization schemes $\mathcal{L}_i, \mathcal{L}_f$ on a space $\Omega$. Let $\nu$ be a measure on $\Omega$ and consider localization the process $\left(\nu_t^{(i)}\right)_t$ obtained from applying $\mathcal{L}_i$ to $\nu$. Let $\mathfrak{a}$ be a stopping time adapted to the filtration of this process. Now, conditional on $\nu_{\mathfrak{a}}^{(i)}$, let $\left(\nu_t^{(f)}\right)_t$ be a localization process obtained from applying $\mathcal{L}_f$ to $\nu_{\mathfrak{a}}^{(i)}$. We may obtain a localization process $(\nu_t)_t$ by defining

$$\nu_t := \begin{cases} \nu_t^{(i)} & t \leq \mathfrak{a} \\ \nu_{t-\tau}^{(f)} & t \geq \mathfrak{a}. \end{cases}$$

We denote the localization scheme mapping $\nu$ to $(\nu_t)_t$ by $\mathrm{concat}(\mathcal{L}_i, \mathcal{L}_f, \mathfrak{a})$.

The advantage of concatenating two localization schemes is demonstrated by the following theorem.

**Theorem 46.** *(Annealing - variance). Let $\mathcal{L}_i, \mathcal{L}_f$ be two localization schemes on $\Omega$ such that the latter is a Doob localization scheme. Let $\nu$ be a probability measure to which we assign a localization process $(\nu_t)_t$ from $\mathrm{concat}(\mathcal{L}_i, \mathcal{L}_f, \mathfrak{a})$ for a stopping time $\mathfrak{a}$. Fix $\tau > 0$ and let $P = P^{(\mathcal{L}_f, \tau)}(\nu_{\mathfrak{a}})$ be the (random) transition kernel associated to $\nu_{\mathfrak{a}}$ via $(\mathcal{L}_f, T)$ using (1). Suppose that,*

1. *We have the approximate variance conservation bound*

$$\frac{\mathbb{E}\left[\mathrm{Var}_{\nu_{\mathfrak{a}}}[\varphi]\right]}{\mathrm{Var}_{\nu}[\varphi]} \geq \varepsilon, \quad \forall \varphi : \Omega \to \mathbb{R}.$$

2. *Almost-surely, we have*

$$\mathfrak{gap}(P(\nu_{\mathfrak{a}})) \geq \delta.$$

*Under these two assumptions, we have that the Markov chain associated to $\nu$ via $\mathcal{L}_f$ has spectral gap at least $\varepsilon\delta$.*

In the above theorem, the localization scheme $\mathcal{L}_i$ is thought of as an annealing procedure which takes a (potentially not well-behaved) measure $\nu$, and outputs a measure $\nu_{\mathfrak{a}}$ which is well-behaved in the sense of assumption 2. Assumption 1 tells us that we did not lose much of the variance throughout the annealing process. The proof relies on the fact that the Dirichlet form associated by a Doob localization is a supermartingale - this fact was crucially used in [EKZ21] in the context of stochastic localization.

Next, we have a completely analogous theorem for the entropy.

**Theorem 47.** *(Annealing - entropy). Let $\mathcal{L}_i, \mathcal{L}_f$ be two localization schemes on $\Omega$ such that the latter is a Doob localization scheme. Let $\nu$ be a probability measure to which we assign a localization process $(\nu_t)_t$ from $\mathrm{concat}(\mathcal{L}_i, \mathcal{L}_f, \mathfrak{a})$ for a stopping time $\mathfrak{a}$. Fix $\tau > 0$ and let $P = P^{(\mathcal{L}_f, \tau)}(\nu_{\mathfrak{a}})$ be the (random) transition kernel associated to $\nu_{\mathfrak{a}}$ via $(\mathcal{L}_f, T)$ using (1). Suppose that,*

1. *We have the entropy conservation inequality*

$$\frac{\mathbb{E}\left[\mathrm{Ent}_{\nu_{\mathfrak{a}}}[f]\right]}{\mathrm{Ent}_{\nu}[f]} \geq \varepsilon, \quad \forall f : \Omega \to \mathbb{R}_+.$$

2. *Almost-surely, one has*

$$\rho_{\mathrm{LS}}(P(\nu_{\mathfrak{a}})) \geq \delta. \tag{34}$$

*Under these two assumptions, we have that the Markov chain associated to $\nu$ via $\mathcal{L}_f$ has an MLSI coefficient of at least $\varepsilon\delta$.*

Let $P^{(t)} = P^{(\mathcal{L}_f,\tau)}(\nu_t)$. Our main step towards proving the two theorems is the following.

**Proposition 48.** *For every $\varphi : \Omega \to \mathbb{R}$ and $f : \Omega \to \mathbb{R}_+$, the processes*

$$t \to \int_{\Omega\times\Omega} \varphi(x)\varphi(y) dP_x^{(t)}(y) d\nu_t^{(i)}(x)$$

*and*

$$t \to \int_{\Omega} P^{(t)} f(x) \log P^{(t)} f(x) d\nu_t^{(i)}(x)$$

*are submartingales.*

The proof of Proposition 48 is deferred to Appendix A.1. Let us see how theorems 46 and 47 follow from this proposition.

*Proof of Theorem 46.* We write

$$\begin{aligned}
&\frac{1}{2}\int_\Omega\int_\Omega (\varphi(x)-\varphi(y))^2 dP_x^{(0)}(y) d\nu(x) \\
&= \int_\Omega \varphi(x)^2 d\nu(x) - \int_\Omega\int_\Omega \varphi(x)\varphi(y) dP_x^{(0)}(y) d\nu(x) \\
&\overset{(i)}{\geq} \mathbb{E}\left[\int_\Omega \varphi(x)^2 d\nu_{\mathfrak{a}}^{(i)}(x) - \int_\Omega\int_\Omega \varphi(x)\varphi(y) dP_x^{(\mathfrak{a})}(y) d\nu(x)\right] \\
&= \mathbb{E}\left[\frac{1}{2}\int_\Omega\int_\Omega (\varphi(x)-\varphi(y))^2 dP_x^{(\mathfrak{a})}(y) d\nu_{\mathfrak{a}}^{(i)}(x)\right] \\
&\overset{(ii)}{\geq} \delta\mathbb{E}\left[\mathrm{Var}_{\nu_{\mathfrak{a}}^{(i)}}[\varphi]\right] \\
&\overset{(iii)}{\geq} \delta\varepsilon \mathrm{Var}_\nu[\varphi],
\end{aligned}$$

where (i) uses Proposition 48 and the optional stopping theorem, (ii) uses the second assumption of the theorem and (iii) uses the first assumption of the theorem. ∎

*Proof of Theorem 47.* We have,

$$\begin{aligned}
\mathrm{Ent}_\nu[f] - \mathrm{Ent}_\nu[P^{(0)}f] &= \int_\Omega f(x)\log f(x) d\nu(x) - \int_\Omega P^{(0)}f(x)\log P^{(0)}f(x) d\nu(x) \\
&\overset{(i)}{\geq} \mathbb{E}\left[\int_\Omega f(x)\log f(x) d\nu_{\mathfrak{a}}^{(i)}(x) - \int_\Omega P^{(\mathfrak{a})}f(x)\log P^{(\mathfrak{a})}f(x) d\nu_{\mathfrak{a}}^{(i)}(x)\right] \\
&= \mathbb{E}\left[\mathrm{Ent}_{\nu_{\mathfrak{a}}^{(i)}}[f] - \mathrm{Ent}_{\nu_{\mathfrak{a}}^{(i)}}[P^{(\mathfrak{a})}f]\right] \\
&\overset{(ii)}{\geq} \delta\mathbb{E}\left[\mathrm{Ent}_{\nu_{\mathfrak{a}}^{(i)}}[f]\right] \\
&\overset{(iii)}{\geq} \delta\varepsilon\mathrm{Ent}_\nu[f],
\end{aligned}$$

where (i) uses Proposition 48 and the optional stopping theorem, (ii) uses the second assumption of the theorem and (iii) uses the first assumption of the theorem. □

# 5 Applications

In this section we provide several applications to demonstrate ways to obtain mixing time results via our framework.

## 5.1 Glauber dynamics for Ising models

As a first example, we will prove a general theorem that provides a sufficient condition for MLSI, which as a special case recovers two results: the first one was shown in [EKZ21, AJK+21b] and establishes mixing time for Ising models whose interaction matrix has an operator norm bounded by a constant and in particular is relevant to the **Sherrington-Kirkpatrick** model in high-enough temperatures. The second one concerns Ising models in the uniqueness regime and improves upon the main theorem of [CFYZ21a].

Both applications will rely on the following result, which is obtained through a simple combination of several ingredients of our framework.

**Theorem 49.** *Let $\nu$ be a measure on $\{-1,1\}^n$ and let $J$ be a positive-definite $n \times n$ matrix. For any $0 \leq \lambda \leq 1$ and every $v \in \mathbb{R}^n$, consider the probability measure $\mu_{\lambda,v}$ defined by*

$$\frac{d\mu_{\lambda,v}}{d\nu}(x) \propto \exp\left(-\lambda\langle x, Jx\rangle + \langle v, x\rangle\right). \tag{35}$$

*Suppose that, for some $\alpha : [0,1] \to \mathbb{R}_+$,*

$$\|\mathrm{Cov}(\mu_{\lambda,v})\|_{\mathrm{OP}} \leq \alpha(\lambda), \quad \forall \lambda \in [0,1],\ v \in \mathbb{R}^n, \tag{36}$$

*and that for some $\varepsilon > 0$,*

$$\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\mu_{1,v})) \geq \varepsilon, \quad \forall v \in \mathbb{R}^n, \tag{37}$$

*where $P^{\mathrm{GD}}(\cdot)$ denotes the transition kernel of the Glauber dynamics. Then,*

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

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

$$\nu_{J,v}(\{x\}) \propto \exp\left(\langle x, Jx\rangle + \langle x, v\rangle\right), \tag{38}$$

referred to as the Ising measure with interaction matrix $J$ and external field $v$.

If we apply the above theorem to the measure $\nu = \nu_{J,v}$ (and with $J$ taken to be the interaction matrix), we have that the measure $\mu_{1,v}$ defined in (35) is just a product measure, so that condition (37) is trivially satisfied with $\varepsilon = \frac{1}{n}$. In order to successfully apply the theorem, it therefore remains to verify condition (36). We will show how to do so in two different settings.

*Proof of Theorem 49.* Consider the localization process $(\nu_t)_t$ corresponding to $\nu$ obtained via the stochastic localization (5), with the choice $C_t = (2J)^{1/2}$, up to time $t = 1$. Consider the transition kernel $P^{\mathrm{GD}}(\nu_1)$, hence, the random transition kernel associated to the measure $\nu_1$ via the Glauber dynamics. We apply Theorem 47 in order to obtain a lower bound on $\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu))$.

According to equation (6), $\nu_t$ 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 $y_t$, meaning that $\nu_t = \mu_{t,y_t}$. Since in particular, $\nu_1 = \mu_{1,y_1}$, the assumption (37) then amounts to the fact that $\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu_1)) \geq \varepsilon$, which fulfills the second condition of Theorem 47.

Next, the assumption (36) tells us that

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

By applying Lemma 40 with the choice $A = \alpha(t)\mathbf{I}_n$ and $C = (2J)^{1/2}$, we conclude that $\nu_t$ is $\alpha$-entropically stable with respect to $\psi(x,y) = \frac{1}{2}\left|(2J)^{1/2}(x-y)\right|^2$ with $\alpha = 2\|J\|_{\mathrm{OP}}\alpha(t)$. Invoking Proposition 39 gives the approximate-entropy-conservation bound

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

Therefore, condition 1 of Theorem 47 is satisfied with $\alpha = e^{-2\|J\|_{\mathrm{OP}}\int_0^1 \alpha(t)dt}$, and we conclude that $\rho_{\mathrm{LS}}(P) \geq \varepsilon e^{-2\|J\|_{\mathrm{OP}}\int_0^1 \alpha(t)dt}$. This completes the proof. □

### 5.1.1 Mixing under a spectral condition on the interaction matrix

It turns out that if the interaction matrix has operator norm bounded by a certain constant, then the covariance matrix of the model is also bounded. The following lemma is based on the decomposition obtained by Bauerschmidt and Bodineau [BB19]. For completeness, we include a proof in Appendix B.

**Lemma 50.** *Let $J$ be a positive-definite matrix with $\|J\|_{\mathrm{OP}} < \frac{1}{2}$ and let $v \in \mathbb{R}^n$. One has*

$$\|\mathrm{Cov}(\nu_{J,v})\|_{\mathrm{OP}} \leq \frac{1}{1-2\|J\|_{OP}}.$$

Combining the lemma with Theorem 49, we obtain

**Corollary 51.** *Let $J$ be a positive-definite matrix with $\|J\|_{\mathrm{OP}} < \frac{1}{2}$ and let $v \in \mathbb{R}^n$. The Glauber dynamics for $\nu = \nu_{J,v}$ mixes in time $O(n\log n)$.*

*Proof.* An application of Lemma 50 gives

$$\|\mathrm{Cov}(\mu_{\lambda,v})\|_{\mathrm{OP}} \preceq \frac{1}{1-2(1-\lambda)\|J\|_{\mathrm{OP}}},$$

so the assumption (36) holds with $\alpha(\lambda)$ equal to the right-hand side. Assumption (37) holds with $\delta = 1/n$ since $\mu_{1,v}$ is a product measure. The conclusion of the Theorem 49 tells us that

$$\begin{aligned}\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu_{J,v})) &\geq \frac{1}{n}\exp\left(-\int_0^1 \frac{2}{1/\|J\|_{\mathrm{OP}} - 2(1-\lambda)}d\lambda\right)\\ &= \frac{1}{n}(1-2\|J\|_{\mathrm{OP}}).\end{aligned}$$

The proof is complete via fact 2. □

*Remark* 52. The above corollary implies an optimal mixing bound for Glauber dynamics on the Sherrington-Kirkpatrick model up to inverse temperature $\beta = 1/4$ (see [EKZ21] for a detailed discussion on the Sherrington-Kirkpatrick model). Very recently, El Alaoui, Montanari and Sellke [AMS22] gave a polynomial time sampling algorithm which is valid all the way to the critical temperature $\beta = 1/2$. The idea behind their algorithm is to simulate the stochastic localization process. To do so, one needs to estimate the center of mass of tilted measures which, in turn, is done using approximate message passing.

### 5.1.2 Near-critical ferromagnetic Ising models

Next, we show how to recover a variant of the result which appears in the very recent paper of Bauerschmidt and Dagallier [BD22]. Consider a measure of the form (38) which is *ferromagnetic* in the sense that $J_{i,j} \geq 0$ for all $i \neq j$. For every $\lambda \in (0,1)$ define

$$\chi(\lambda) := \|\mathrm{Cov}(\nu_{\lambda J,0})\|_{\mathrm{OP}}$$

In the ferromagnetic case, following theorem, due to Ding, Song and Sun, is very helpful towards verifying assumption (36) in Theorem 49.

**Theorem 53.** *([DSS22]) For any interaction matrix $J$ which is ferromagnetic (hence $J_{i,j} \geq 0$ for all $i \neq j$), for all $i \neq j$ and for all $v \in \mathbb{R}^n$, we have*

$$\mathrm{Cov}(\nu_{J,v})_{i,j} \leq \mathrm{Cov}(\nu_{J,0})_{i,j}.$$

By the above combined with the Perron-Frobenius theorem, it follows that

$$\|\mathrm{Cov}(\nu_{\lambda J,v})\|_{\mathrm{OP}} \leq \chi(\lambda), \quad \forall v \in \mathbb{R}^n.$$

Therefore, invoking Theorem 49 with $\mu_{\lambda,v} = \nu_{(1-\lambda)J,v}$ immediately yields the following.

**Corollary 54.** *For any ferromagnetic interaction matrix $J$ which is positive definite, we have*

$$\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu_{J,v}) \geq \frac{1}{n} \exp\left(-2\|J\|_{\mathrm{OP}} \int_0^1 \chi(\lambda) d\lambda\right).$$

As noted in [BD22], the positive definiteness can always be imposed by adding a diagonal matrix without changing the corresponding Ising model.

### 5.1.3 Graphical Ising model in the uniqueness regime

For a graph $G = (V, E)$ with $V = [n]$ and $\beta \in \mathbb{R}$, the corresponding Ising model is defined as

$$\nu_{G,\beta}(x) \propto \exp\left(\langle x, v\rangle + \beta \sum_{(i,j)\in E} \mathbf{1}_{\{x_i = x_j\}}\right). \tag{39}$$

We say that the measure $\nu_{G,\beta}$ satisfies the tree-uniqueness condition if

$$\exp(|\beta|) < \frac{\Delta(G) - \delta}{\Delta(G) - 2 + \delta}, \tag{40}$$

for $\delta \in (0,1)$, where $\Delta(G) \geq 3$ is the maximum degree of the graph. The following result was proven in [CLV21] (see also [CFYZ21a, Lemma 8.3]).

**Theorem 55.** *(Spectral independence for the Ising model). Let $\nu = \nu_{G,\beta}$ be an Ising model defined as in (39) which satisfies the tree-uniqueness condition (40). Then for any external field $v \in \mathbb{R}^n$,*

$$\rho(\mathbf{\Psi}(\mathcal{T}_v \nu)) \leq \frac{2}{\delta}.$$

Given a graph $G = (V, E)$, define a matrix $J = J_G$ by

$$J_{i,j} = \frac{1}{2}\left(\mathbf{1}_{\{(i,j)\in E)\}} + \mathbf{1}_{\{i=j\}}\deg_G(i)\right).$$

It is straightforward to check that

$$\sum_{(i,j)\in E} \mathbf{1}_{\{x_i = x_j\}} = \langle x, J_G x\rangle,$$

and that $J_G$ is positive-definite.

**Corollary 56.** *Any graphical Ising model $\nu$ of the form (39) which satisfies (40), one has*

$$\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu)) \geq \frac{\exp(-8/\delta)}{n}.$$

*Consequently, the mixing time for Glauber dynamics is $O(n \log n)$.*

*Proof.* First, we deal with the case $\beta > 0$. Apply Theorem 49 with the choice $J = \beta J_G$. Observe that for all $\lambda \in [0, 1]$, the measure $\mu_{\lambda,v}$ defined in (35) is of the form (39), with a smaller value of $\beta$, and therefore $\mu_{\lambda,v}$ is also in the uniqueness regime. An application of Theorem 55 ensures that $\rho(\mathbf{\Psi}(\mu_{\lambda,v})) \leq \frac{2}{\delta}$. By Fact 23 we have

$$\|\mathrm{Cov}(\mu_{\lambda,v})\|_{\mathrm{OP}} \leq \|\mathbf{Cor}(\mu_{\lambda,v})\|_{\mathrm{OP}} \leq \frac{2}{\delta},$$

so (36) holds with $\alpha(\lambda) = 2/\delta$. Moreover, since $\mu_{1,v}$ is a product measure, condition (37) holds with $\varepsilon = 1/n$. Therefore, we can conclude by Theorem 49 that

$$\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu)) \geq \frac{1}{n}\exp\left(-4\frac{\beta\|J_G\|_{\mathrm{OP}}}{\delta}\right) \geq \frac{1}{n}e^{-8/\delta}.$$

The last step follows because $J_G = \mathbf{I}_n - \frac{1}{2}L^{\mathrm{sym}}$ where $L^{\mathrm{sym}}$ is the symmetrically normalized Laplacian which has eigenvalues between 0 and 2; and $\beta < \log\left(\frac{\Delta(G)-\delta}{\Delta(G)-2+\delta}\right) \leq 2$.
Second, the case $\beta = 0$ is trivial because it becomes a product measure. Finally, for the case $\beta < 0$, we observe that

$$-|\beta|\langle x, J_G x\rangle = |\beta|\left(\langle x, (\mathbf{I}_n - J_G)x\rangle - \langle x, x\rangle\right) == |\beta|\left(\langle x, (\mathbf{I}_n - J_G)x\rangle - n\right).$$

$\mathbf{I}_n - J_G$ is positive-definite and has operator norm bounded by 1. Using the same argument as the case $\beta > 0$ but with $J_G$ replaced by $\mathbf{I}_n - J_G$, we obtain the same MLSI lower bound. ∎

## 5.2 Sampling from strongly log-concave measures via a Restricted Gaussian Oracle

Let $\nu$ be a log-concave measure on $\mathbb{R}^n$, hence, a measure of the form

$$\frac{\nu(dx)}{dx} = \exp(-V(x)), \quad V : \mathbb{R}^n \to \mathbb{R} \text{ convex}.$$

We assume that it is uniformly (or strongly) log-concave, namely that

$$\nabla^2 V(x) \succeq \mu \mathbf{I}_n, \quad \forall x \in \mathbb{R}^n. \tag{41}$$

Let $(\nu_t)_t$ be the process attained from $\nu$ via the stochastic localization scheme (5) with the choice $C_t \equiv \mathbf{I}_n$. According to equation (6), $\nu_t$ has the form

$$\frac{d\nu_t}{dx} \propto \exp\left(-\tilde{V}(x) - \frac{t}{2}|x|^2\right) \tag{42}$$

where $\tilde{V}(x)$ is convex (it is an exponential tilt of $V(x)$). The following well-known bound goes back to Brascamp and Lieb.

**Theorem 57.** *If $\rho$ is a measure on $\mathbb{R}^n$ of the form $\frac{\rho(dx)}{dx} \propto \exp(-U(x))$ that satisfies the uniform convexity condition*

$$\nabla^2 U(x) \succeq \alpha \mathbf{I}_n, \quad \forall x \in \mathbb{R}^n,$$

*then $\|\mathrm{Cov}(\rho)\|_{\mathrm{OP}} \leq \frac{1}{\alpha}$.*

Equations (41) and (42) imply that $\nu_t$ has the form $\frac{d\nu_t(x)}{dx} = \exp(-U(x))$ with

$$\nabla^2 U(x) \succeq (\mu + t)\mathbf{I}_n, \quad \forall x \in \mathbb{R}^n.$$

Therefore, the above theorem gives

$$\|\mathrm{Cov}(\mathcal{T}_v \nu_t)\|_{\mathrm{OP}} \leq \frac{1}{\mu + t}, \quad \forall v \in \mathbb{R}^n. \tag{43}$$

An application of Lemma 40 gives that $\nu_t$ is $\alpha$-entropically-stable with respect to $\psi(x, y) = \frac{1}{2}|x - y|^2$, with $\alpha = \frac{1}{\mu+t}$. Proposition 39 now gives the approximate conservation of entropy bound

$$\mathbb{E}\left[\mathrm{Ent}_{\nu_T}[f]\right] \geq \mathrm{Ent}_{\nu}[f] \exp\left(-\int_0^T \frac{1}{\mu + t} dt\right) = \frac{\mu}{\mu + T}\mathrm{Ent}_{\nu}[f],$$

for an arbitrary $f : \mathbb{R}^n \to \mathbb{R}^+$ such that $\int f d\nu > 0$. Using equation (11), this gives

$$\rho_{\mathrm{LS}}(P^{\mathrm{RGD}_\eta}(\nu)) \geq \frac{\mu}{\mu + 1/\eta},$$

where $P^{\mathbf{RGD}_\eta}(\nu)$ is the transition kernel of the restricted Gaussian dynamics (Definition 15 above). We have proved:

**Theorem 58.** *If $\nu$ is a $\mu$-strongly log-concave measure on $\mathbb{R}^n$, then the associated restricted Gaussian dynamics with parameter $\eta$ satisfies $\rho_{\mathrm{LS}}(P^{\mathrm{RGD}_\eta}(\nu)) \geq \frac{\mu}{\mu+1/\eta}$.*

The above theorem not only recovers Theorem 1.1 in [LST21], but also resolves the open problem discussed in section 1.4 of the same paper: it provides a proof to show that the Restricted Gaussian Dynamics mixes in $\eta\mu \log(n)$ steps when started from any $O(\exp(n))$-warm start in the case $\eta < 1/\mu$. As defined in [LST21], an initial measure $\rho$ is a $\beta$-warm start with respect to $\nu$ if $\frac{d\rho}{d\nu}(x) \leq \beta$ everywhere. The applications of the above theorem includes improving dependency of condition number dependency of Markov chains for log-concave sampling and introducing new sampling algorithms for composite log-concave distributions and log-concave finite sums, which has been extensively discussed in [LST21].

## 5.3 A spectral-independence based condition for fast mixing and applications to sampling from the hardcore model

In this section we present the main application of the negative-fields localization, giving a sufficient condition for fast mixing. In particular, it provides an optimal mixing time bound for sampling the hardcore model via Glauber dynamics.

Recall that, for $u \in \{-1, 0, 1\}$, we define the $u$-pinning $\mathcal{R}_u\nu$ to be the restriction of $u$ to the subcube $S_u := \{x \in \{-1, 1\}^n;\ x_i u_i \geq 0,\ \forall i \in [n]\}$ and that $\mathcal{T}_v\nu$ is the $v$-exponential tilt of $\nu$.

The following definition relates the marginals of the pinned distribution to those of the original distribution and lower bounds the marginals in a one-sided fashion.

**Definition 59.** For $K > 1$, a measure $\nu$ on $\{-1, 1\}^n$ has $K$-tame marginals if for every $i \in [n]$ and for every $u, w \in \{-1, 0, 1\}^n$ with $\mathrm{supp}(u) \cap \mathrm{supp}(w) = \emptyset$ with $u_i = w_i = 0$, we have

$$\frac{1 + \mathbf{b}_i(\mathcal{R}_w\mathcal{R}_u\nu)}{1 - \mathbf{b}_i(\mathcal{R}_w\mathcal{R}_u\nu)} \leq K\frac{1 + \mathbf{b}_i(\mathcal{R}_u\nu)}{1 - \mathbf{b}_i(\mathcal{R}_u\nu)}, \text{ and} \tag{44}$$

$$1 - \mathbf{b}_i(\mathcal{R}_u\nu) \geq \frac{1}{K}. \tag{45}$$

Our main aim is to establish the following sufficient condition for a MLSI: Given a measure $\nu$ on $\{-1, 1\}^n$, if a MLSI can be established for a perturbed measured obtained after applying an external field $v$, and if the correlation matrix has bounded operator norm for any external field $\lambda v$, $\lambda \in [0, 1]$ and under all pinnings, then the original measure satisfies a MLSI.

**Theorem 60.** *Let $s > 0$. Let $\nu$ be a measure on $\{-1, 1\}^n$ such that for any $\lambda \in [0, s]$, $\mathcal{T}_{-\lambda\vec{\mathbf{1}}}\nu$ has $K$-tame marginals. Define $\vec{\mathbf{1}} := (1, ..., 1)$. Suppose that for any $\lambda \in [0, s]$ and $u \in \{-1, 0, 1\}^n$, we have*

$$\left\|\mathbf{Cor}(\mathcal{R}_u\mathcal{T}_{-\lambda\vec{\mathbf{1}}}\nu)\right\|_{\mathrm{OP}} \leq \eta. \tag{46}$$

*Moreover, assume that for all $u \in \{-1, 0, 1\}^n$,*

$$\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\mathcal{R}_u\mathcal{T}_{-s\vec{\mathbf{1}}}\nu)) \geq \delta. \tag{47}$$

*Then we have*

$$\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu)) \geq \delta e^{-cK^4\eta},$$

*for a universal constant $c > 0$.*

*Remark* 61. In the above theorem we considered the "path" of tilts $\{-\lambda\vec{\mathbf{1}}; \lambda \in [0, s]\}$. The same proof works for the more general case that we consider $\mathcal{T}_{v(\lambda)}\nu$ for an arbitrary curve $v : [0, 1] \to \mathbb{R}^n$. The special case $v(\lambda) = -\lambda\vec{\mathbf{1}}$ is sufficient for the applications that we know of, which is why we chose to stick to it for the sake of simplicity.

*Proof of Theorem 60.* The main proof strategy is to use the negative-fields-localization (constructed in Section 2.4.4) as an annealing scheme for the measure. Namely, we apply Proposition 18 with the choice $v(t) = -t\vec{\mathbf{1}}$ to obtain a localization process $(\nu_t)_t$. We concatenate this process to the coordinate-by-coordinate localization scheme at time $s$ and apply theorem 47, which gives us a MLSI bound for the Glauber dynamics $P^{\mathrm{GD}}(\nu)$.

According to Proposition 18, for every $t > 0$ there exists $u(t) \in \{-1, 0, 1\}^n$ such that

$$\nu_t = \mathcal{T}_{-t\vec{\mathbf{1}}}\mathcal{R}_{u(t)}\nu.$$

Therefore, condition 2 of theorem 47 follows immediately from equation (47).

It remains to show that condition 1 of Theorem 47 is satisfied, which boils down to the approximate entropy conservation bound

$$\mathbb{E}[\mathrm{Ent}_{\nu_s}[f]] \geq e^{-64K^2e^s\eta}\mathrm{Ent}_{\nu}[f], \tag{48}$$

for an arbitrary function $f : \{-1,1\}^n \to \mathbb{R}_+$.

Fix $t \in [0,s]$. We aim to apply Theorem 42 to the measure $\nu_t$, which should be understood as a measure on the sub-cube $S_{u(t)} = \{x \in \{-1,1\}^n;\ x_i u(t)_i \geq 0,\ \forall i \in [n]\}$. To check that condition (30) is fulfilled, let $\tilde{u} \in \{-1,0,1\}^n$ be such that $\tilde{u}_i u(t)_i = 0$ for all $i \in [n]$ (which amounts to all valid pinning for a function on $S_{u(t)}$), then

$$\begin{aligned}\|\mathbf{Cor}(\mathcal{R}_{\tilde{u}}\nu_t)\|_{\mathrm{OP}} &= \left\|\mathbf{Cor}(\mathcal{R}_{\tilde{u}}\mathcal{R}_{u(t)}\mathcal{T}_{-t\vec{\mathbf{1}}}\nu)\right\|_{\mathrm{OP}} \\ &= \left\|\mathbf{Cor}(\mathcal{R}_{\tilde{u}+u(t)}\mathcal{T}_{-t\vec{\mathbf{1}}}\nu)\right\|_{\mathrm{OP}} \overset{(46)}{\leq} \eta.\end{aligned}$$

Define $F_t = \{i \in [n]; u_i(t) = 0\}$, the set of coordinates which have not been pinned yet. In order to verify condition (31) and (32) for $\nu_t$, we use the fact that $\nu_t$ is obtained by pinning $\mathcal{T}_{-t\vec{\mathbf{1}}}\nu$ which has $K$-tame martingales. First, for all $i \in F_t$ and all $u(t) \perp \tilde{u} \in \{-1,0,1\}^n$ with $\tilde{u}_i = 0$, $1 - \mathbf{b}_i(\mathcal{R}_{\tilde{u}}\nu_t) = 1 - \mathbf{b}_i(\mathcal{R}_{\tilde{u}}\mathcal{R}_{u(t)}\mathcal{T}_{-t\vec{\mathbf{1}}}\nu) \geq \frac{1}{K}$, which verifies condition (32) for $\nu_t$. Second, using the $K$-tame marginals of $\mathcal{T}_{-t\vec{\mathbf{1}}}\nu$, we obtain

$$\begin{aligned}\frac{1+\mathbf{b}_i(\mathcal{R}_w\mathcal{R}_{\tilde{u}}\nu_t)}{1-\mathbf{b}_i(\mathcal{R}_w\mathcal{R}_{\tilde{u}}\nu_t)} &= \frac{1+\mathbf{b}_i(\mathcal{R}_w\mathcal{R}_{\tilde{u}}\mathcal{R}_{u(t)}\mathcal{T}_{-t\vec{\mathbf{1}}}\nu)}{1-\mathbf{b}_i(\mathcal{R}_w\mathcal{R}_{\tilde{u}}\mathcal{R}_{u(t)}\mathcal{T}_{-t\vec{\mathbf{1}}}\nu)} \\ &\leq K\frac{1+\mathbf{b}_i(\mathcal{R}_w\nu_t)}{1-\mathbf{b}_i(\mathcal{R}_w\nu_t)},\end{aligned}$$

which verifies condition (31) for $\nu_t$. Apply Theorem 42 to $\nu_t$, we obtain that $\nu_t$ is $cK^4\eta$-entropically stable with respect to the function $H(x,y)$, where $c$ is a universal constant. We can finally use Proposition 41 which gives that for $t, h > 0$,

$$\mathbb{E}[\mathrm{Ent}_{\nu_{t+h}}[f]|u(t)] \geq \mathrm{Ent}_{\nu_t}[f](1 - cK^4\eta h) + o(h).$$

By integrating this inequality, we finally get

$$\mathbb{E}[\mathrm{Ent}_{\nu_s}[f]] \geq \mathrm{Ent}_{\nu}[f]\exp\left(-cK^4\eta\right).$$

Both conditions of Theorem 47 are satisfied, and we conclude the proof. □

### 5.3.1 Application to an optimal mixing bound for Glauber dynamics in the hardcore model

Given a graph $G = (V,E)$, we define the set of all independent assignments as

$$\mathcal{I}_G := \left\{x \in \{-1,1\}^V :\ x_i + x_j \leq 1,\ \forall (i,j) \in E\right\}.$$

If we associate a point $x \in \{-1,1\}^V$ the set $A(x) = \{v \in V; x_v = 1\}$, then $A(\mathcal{I}_G)$ is the family of independent sets in $G$.

For $\lambda \in (0,\infty)$ let $\rho_\lambda$ be the product probability measure on $\{-1,1\}^V$ defined by

$$\rho_\lambda(\{x\}) = \prod_{v\in V}\frac{\mathbf{1}_{\{x_v=-1\}} + \lambda\mathbf{1}_{\{x_v=1\}}}{1+\lambda}.$$

The *hardcore* model on $G$ with *fugacity* $\lambda$ is a probability measure on $\{-1,1\}^V$ defined by

$$\nu_{G,\lambda}(\{x\}) := \frac{\rho_\lambda(\{x\})\mathbf{1}_{\{x\in\mathcal{I}_G\}}}{\rho_\lambda(\mathcal{I}_G)}.$$

A hardcore model with fugacity $\lambda$ is called $\delta$-**unique** if $\lambda \leq (1-\delta)\lambda_\Delta$, where the critical fugacity is

$$\lambda_\Delta = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}, \tag{49}$$

and where $\Delta$ is the maximal degree of $G$.

Our goal is to prove the following theorem, which asserts that the Glauber dynamics mixes in time $O(n\log n)$ in the uniqueness regime.

**Theorem 62.** *Given a graph $G=(V,E)$. $|V|=n$ and $\lambda>0$, let $\nu=\nu_{G,\lambda}$ be the hardcore model on $G$ with fugacity $\lambda$. Suppose that $\nu$ is $\delta$-unique for some $\delta>0$. Let $\mu_0$ be an arbitrary initial distribution supported on $\mathcal{I}_G$. Then there exists a universal constant $c>0$, such that*

$$t_{\mathrm{mix}}(P^{\mathrm{GD}}(\nu),\varepsilon;\mu_0) \leq \exp\left(\frac{c}{\delta}\right)\left(n\log(n)+3n\log(1/\varepsilon)\right).$$

Proving mixing bounds for the hardcore model has been one of the central applications of the spectral and entropic independence frameworks. This model has inspired the first paper which put forth the notion of spectral independence as well as many of the following works. Let us summarize the progress made so far.

- Anari, Liu and Oveis Gharan were the first to introduce the notion of spectral independence in [ALOG20], and proved an $n^{\exp(O(1/\delta))}$ mixing time.
- Chen, Liu and Vigoda [CLV21] proved tight bounds on spectral independence, gave an improved $n^{O(1/\delta)}$ mixing time.
- Chen, Liu and Vigoda [CLV20] extended the framework towards MLSI bounds and proved that $\Delta^{O(\Delta^2/\delta)}n\log(n)$, which in particular gives an optimal bound for constant $\Delta$.
- Jain, Pham and Vuong [AJK+21a] proved a mixing time of $\Delta^{O(1/\delta)}n^2$.
- Chen, Feng, Yin and Zhang [CFYZ21b], in Theorem 1.3, obtained an optimal bound for the relaxation time, which implies a mixing time of $C(\delta)n^2\log(\Delta)$.
- Anari et al. [AJK+21a, Theorem 1], showed that a variant of Glauber dynamics called the balanced Glauber dynamics gives an $O_\delta(n\log(n))$ mixing time.

Our result, Theorem 62 is the first to give an optimal mixing bound for the (usual) Glauber dynamics.

The proof combines many of ingredients developed in our framework, together with the idea of using a certain "restricted" type of entropy decay.

### 5.3.2 Properties of the hardcore distribution

In this subsection, we collect four properties of the hardcore distribution: spectral independence for all negative tilts and all pinnings, marginal upper and lower bounds, the fact that exponentially tilt hardcore model remains a hardcore model and MLSI coefficient for hardcore model with small fugacity. Throughout this section we fix graph $G = (V, E)$ with maximum degree $\Delta$, a fugacity $\lambda \in (0, \infty)$ and define $\nu = \nu_{G,\lambda}$ to be the corresponding hardcore model. Without loss of generality, we identify $V$ with $[n]$ for simplicity.

The first lemma establishes spectral independence of all negative tilts and all pinnings of the hardcore distribution, and was proved in Lemma 8.4 of [CFYZ21b] extending the analyses in [CLV20].

**Lemma 63.** *Consider the hardcore model $\nu = \nu_{G,\lambda}$ with $\Delta = \Delta(G) \geq 3$. Furthermore, suppose that $\nu$ is $\delta$-unique on $G$. Then for all $u \in \{-1, 0, 1\}^n$ and for every vector $v \in (-\infty, 0]^V$, we have*

$$\|\mathbf{Cor}(\mathcal{T}_v \mathcal{R}_u \nu)\|_{\mathrm{OP}} \leq \frac{144}{\delta}.$$

The next lemma gives bounds for the marginals of the hardcore distribution. It was essentially proven in [AJK+21a, Proposition 50], but we provide a proof in Appendix C for completeness.

**Lemma 64.** *Let $\nu = \nu_{G,\lambda}$ with $\Delta(G) \geq 3$. Then for any $v \in V$, for any $u \in \{-1, 0, 1\}^n$ which sets all neighbors of $v$ to 0 (i.e., satisfying $u_a = 0, \forall a \in N_v$), we have*

$$\frac{\lambda}{1+\lambda}\left(\frac{1}{1+\lambda}\right)^{|N_v|} \leq \mathbb{P}_{\sigma\sim\nu}\left(\sigma_v = +1 \mid \sigma_i u_i \geq 0, \forall i \in [n]\right) \leq \frac{\lambda}{1+\lambda}.$$

*Moreover, if the model is $\delta$-unique, namely $\lambda \leq (1-\delta)\lambda_\Delta$, then*

$$\frac{\lambda}{1+\lambda} e^{-3e^2} \leq \mathbb{P}_{\sigma\sim\nu}\left(\sigma_v = +1 \mid \sigma_i u_i \geq 0, \forall i \in [n]\right) \leq \frac{\lambda}{1+\lambda}.$$

Recall that $\vec{\mathbf{1}} = (1, \ldots, 1)$. The next lemma shows that the exponentially tilted hardcore model remains a hardcore model.

**Lemma 65.** *For every $G$ and $\lambda$, we have for all $t > 0$,*

$$\mathcal{T}_{-t\vec{\mathbf{1}}} \nu_{G,\lambda} = \nu_{G,e^{-2t}\lambda}. \tag{50}$$

*In other words, the exponential tilt $\mathcal{T}_{-t\vec{\mathbf{1}}}\nu$ corresponds to a hardcore model, on the same graph, with fugacity $e^{-2t}\lambda$.*

*Proof of Lemma 65.* It is clear from the definition of the hardcore model. □

The final lemma states that if the fugacity of the model is small enough, then one has a bound on the MLSI coefficient. This lemma was proved in [EHMT17, Corollary 4.7] (see also [AJK+21a, Proposition 51]).

**Lemma 66.** *Let $\nu$ be a hardcore model on $G$ of maximum degree at most $\Delta$ with fugacity $\lambda \leq \frac{1}{2\Delta}$. $\Delta \geq 3$. Then one has $\rho_{\mathrm{LS}}(P^{GD}(\nu)) \geq \frac{1}{4n}$.*

*Proof.* Combine [AJK+21a, Proposition 51] and [CLV21, Fact 3.5]. □

*Proof of Theorem 62.* We simply verify that all the conditions of Theorem 60 hold. Lemma 63 verifies condition (46) with $\eta = \frac{144}{\delta}$. Next, note that, since $\Delta > 3$,

$$e^{-2t}\lambda \le e^{-2t}\frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} = \frac{e^{-2t}}{\Delta-2}\left(1+\frac{1}{\Delta-2}\right)^{\Delta-1} \le \frac{9e^{1-2t}}{\Delta}.$$

According to Lemma 65, the condition of Lemma 66 is verified for $\mathcal{T}_{-2\vec{\mathbf{1}}}\nu$, which gives that $\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\mathcal{T}_{-2\vec{\mathbf{1}}}\nu)) \ge \frac{1}{4n}$ so that (47) holds true with $s = 2$ and $\delta = 1/(4n)$. Lemma 64 and Lemma 65 together ensure that

$$1 - \mathbf{b}_i(\mathcal{T}_{-t\vec{\mathbf{1}}}\nu) \ge \frac{2}{1+e^{-2t}\lambda} \ge \frac{2}{1+4}.$$

Finally, Lemma 64 ensures that $\mathcal{T}_{-t\vec{\mathbf{1}}}\nu$ has $K$-tame martingales with $K = e^{30}$, for any $t \in [0,s]$. Finally, we invoke Theorem 60 to obtain $\rho_{\mathrm{LS}}(P^{\mathrm{GD}}(\nu) \ge \exp\left(-\frac{c}{\delta}\right)\frac{1}{4n}$ where $c$ is a universal constant. Using Fact 2, this completes the proof. □

# 6 Entropic stability via spectral independence

In this section, we prove Theorem 42, which shows that entropic stability is implied by spectral independence for all pinnings. A priori, the formulation of this theorem has nothing to do with the negative-fields localization, however the main argument of its proof relies on a coupling argument based on the negative-fields localization.

For convenience, we repeat the formulation of Theorem 42 in the following theorem.

**Theorem 67.** *Let $\nu$ be a probability measure on $\{-1,1\}^n$ and let $\alpha \ge 1$. Suppose that*

$$\|\mathbf{Cor}(\mathcal{R}_u\nu)\|_{\mathrm{OP}} \le \alpha, \quad \forall u \in \{-1,0,1\}^n. \tag{51}$$

*Then $\nu$ is $8\alpha$-entropically stable with respect to $\psi(x,y) = \frac{1}{2}|x-y|^2$.*
*Furthermore, if for some constants $K \ge 1, C \ge 1$, for every $i \in [n]$ and for every $u, w \in \{-1,0,1\}^n$ with $\mathrm{supp}(u) \cap \mathrm{supp}(w) = \emptyset$ with $u_i = w_i = 0$, we have*

$$\frac{1+\mathbf{b}_i(\mathcal{R}_w\mathcal{R}_u\nu)}{1-\mathbf{b}_i(\mathcal{R}_w\mathcal{R}_u\nu)} \le K\frac{1+\mathbf{b}_i(\mathcal{R}_u\nu)}{1-\mathbf{b}_i(\mathcal{R}_u\nu)}, \text{ and} \tag{52}$$

$$1 - \mathbf{b}_i(\mathcal{R}_u\nu) \ge \frac{1}{C}, \tag{53}$$

*then $\nu$ is $768\alpha K^3 C$-entropically stable with respect to the function $H(x,y)$ as defined in equation (23).*

The proof of the theorem consists of the following two main lemmas.

**Lemma 68.** *Suppose that $\nu$ satisfies the assumption (51). Then for all $v \in \mathbb{R}^n$,*

$$|\mathbf{b}(\mathcal{T}_v\nu) - \mathbf{b}(\nu)|^2 \le 16\alpha^2|v|^2.$$

*If, furthermore, condition (52) and (53) hold then for all $v \in \mathbb{R}^n$,*

$$\langle v, \mathbf{b}(\mathcal{T}_v\nu) - \mathbf{b}(\nu)\rangle \le 4\alpha\sum_{i\in[n]} K^3 C\,(1+\mathbf{b}_i(\nu))\,v_i^2\exp(4|v_i|).$$

**Lemma 69.** *Let $\nu$ be a measure on $\mathbb{R}^n$. Suppose that*

$$|\mathbf{b}(\mathcal{T}_{v'}\nu) - \mathbf{b}(\nu)|^2 \le \varepsilon^2 |v'|^2, \quad \forall v' \in \mathbb{R}^n, \tag{54}$$

*then*

$$|\mathbf{b}(\mathcal{T}_{v}\nu) - \mathbf{b}(\nu)|^2 \le \varepsilon \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_v\nu||\nu), \quad \forall v \in \mathbb{R}^n.$$

*Furthermore, suppose that for some constant $C \ge 1$, we have*

$$\langle v, \mathbf{b}(\mathcal{T}_v\nu) - \mathbf{b}(\nu)\rangle \le \sum_{i\in[n]} \varepsilon_i v_i^2 \exp(4|v_i|), \quad \forall v \in \mathbb{R}^n, \tag{55}$$

*where $2 \le \varepsilon_i \le C\,(1 + \mathbf{b}_i(\nu))$ for all $i \in [n]$. Then,*

$$H(\mathbf{b}_i(\mathcal{T}_v\nu), \mathbf{b}_i(\nu)) \le 192C \cdot \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_v\nu||\nu), \quad \forall v \in \mathbb{R}^n.$$

Plugging the two lemmas together immediately establishes Theorem 67. Before we move on to the proofs of Lemmas 68 and 69, we need three more technical intermediate lemmas whose proofs are found in Appendix A.

**Lemma 70.** *Let $\nu$ be a probability measure on $\{-1,1\}^n$. If $s \in \{-1,1\}$ and $\mathbf{e}_i$ is a vector of the standard basis, then we have*

$$\mathbf{b}\left(\mathcal{R}_{s\mathbf{e}_i}\nu\right) - \mathbf{b}(\nu) = (1 + s\mathbf{b}(\nu)_i)^{-1}\mathrm{Cov}(\nu)s\mathbf{e}_i. \tag{56}$$

**Lemma 71.** *Define $H(x,y) = \frac{1+x}{2}\log\left(\frac{1+x}{1+y}\right) + \frac{1-x}{2}\log\left(\frac{1-x}{1-y}\right)$ and $\Phi(x) = (1+x)\log(1+x) - x$. Then we have*

$$\frac{1}{2}H(x,y) \le (1+y)\Phi\left(\frac{x-y}{1+y}\right) \le 2H(x,y), \tag{57}$$

*for all $y \in (-1,1)$ and $x \in [-1,1]$. Moreover, for all $\varepsilon \ge 1$,*

$$\frac{1}{4\varepsilon}H(x,y) \le \varepsilon(1+y)\Phi\left(\frac{x-y}{\varepsilon(1+y)}\right). \tag{58}$$

*Finally, we also have*

$$\Phi(|s|) \ge \frac{1}{3}\Phi(s), \quad \forall |s| \le 1. \tag{59}$$

**Lemma 72.** *Let $\rho$ be a measure on $\{-1,1\}^n$ which satisfies, for all $i \in [n]$ and for all $u \in \{-1,0,1\}^n$ such that $u_i = 0$,*

$$1 + \mathbf{b}_i(\mathcal{R}_u\rho) \le \delta, \tag{60}$$

*Then for all $v \in \mathbb{R}^n$ and all $u \in \{-1,0,1\}^n$ with $u_i = 0$, we have*

$$1 + \mathbf{b}_i(\mathcal{T}_v\mathcal{R}_u\rho) \le \delta \exp\left(\max\{0, 2v_i\}\right) \le \delta\exp(2|v_i|). \tag{61}$$

*Moreover, if for all $i \in [n]$ and for all $u \in \{-1,0,1\}^n$ such that $u_i = 0$,*

$$\delta' \le \frac{1 + \mathbf{b}_i(\mathcal{R}_u\rho)}{1 - \mathbf{b}_i(\mathcal{R}_u\rho)} \le \delta'', \tag{62}$$

*Then for all $v \in \mathbb{R}^n$ and all $u \in \{-1,0,1\}^n$ with $u_i = 0$, we have*

$$\delta' \exp\left(\min\{0, 2v_i\}\right) \le \frac{1 + \mathbf{b}_i(\mathcal{T}_v\mathcal{R}_u\rho)}{1 - \mathbf{b}_i(\mathcal{T}_v\mathcal{R}_u\rho)} \le \delta'' \exp\left(\max\{0, 2v_i\}\right). \tag{63}$$

*Proof of lemma 68.* Define $\mu = \mathcal{T}_v \nu$. Consider the localization process $(\mu_t)_t$ obtained via the negative-fields localization starting from $\mu$. More specifically, we invoke Proposition 18 with the choice $v(t) = -tv$, to obtain a process $(u(t))_t$ of pinnings, $u(t) \in \{-1, 0, 1\}^n$, such that the process defined by

$$\mu_t = \mathcal{R}_{u(t)} \mathcal{T}_{v(t)} \mu = \mathcal{R}_{u(t)} \mathcal{T}_{(1-t)v} \nu$$

is a martingale. Moreover, we have

$$\mathbb{P}[u(t+h)_i \neq u(t)_i | u(t)] = h(1 + s_i \mathbf{b}(\mu_s)_i)|v_i| + o(h), \tag{64}$$

where $s_i := \mathrm{sign}(v_i)$.

We couple this process with a process $(\nu_t)_t$ defined by

$$\nu_t = \mathcal{R}_{u(t)} \nu.$$

We remark that $\nu_t$ is **not** a martingale. By definition of $\nu_t$ and $\mu_t$, it is evident that $\nu_1 = \mu_1$ almost surely. Since $\mu_t$ is a martingale, we have

$$\mathbf{b}(\mu) = \mathbb{E}[\mathbf{b}(\mu_1)] = \mathbb{E}[\mathbf{b}(\nu_1)]. \tag{65}$$

Define $F_t = \{i \in [n];\ u(t)_i = 0\}$, the set of coordinates which have not been pinned yet. Note that $\mathbb{P}\big(\|u(t+h) - u(t)\|_1 \geq 2 \ \big|u(t)\big) = o(h)$. Therefore, we have

$$\begin{aligned}
\mathbb{E}[\mathbf{b}(\nu_{t+h}) - \mathbf{b}(\nu_t)|u(t)] &= \mathbb{E}[\mathbf{b}(\mathcal{R}_{u(t+h)}\nu) - \mathbf{b}(\mathcal{R}_{u(t)}\nu)|u(t)] \\
&= \sum_{i \in F_t} \left(\mathbf{b}(\mathcal{R}_{u(t)+s_i e_i}\nu) - \mathbf{b}(\mathcal{R}_{u(t)}\nu)\right) \mathbb{P}(u(t+h)_i \neq u(t)_i) + o(h) \\
&\overset{(64)}{=} \sum_{i \in F_t} \left(\mathbf{b}(\mathcal{R}_{u(t)+s_i e_i}\nu) - \mathbf{b}(\mathcal{R}_{u(t)}\nu)\right) |v_i| h(1 + s_i \mathbf{b}(\mu_t)_i) + o(h) \\
&\overset{(56)}{=} h \sum_{i \in F_t} \mathrm{Cov}(\nu_t) s_i e_i \frac{1 + s_i \mathbf{b}(\mu_t)_i}{1 + s_i \mathbf{b}(\nu_t)_i} |v_i| + o(h).
\end{aligned}$$

Note that if $i \notin F_t$, then $\mathbf{b}(\mu_t)_i \mathbf{b}(\nu_t)_i = 1$ and we can denote by convention that $\frac{0}{0} = 1$. Since $(1 + s_i \mathbf{b}(\nu_t)_i)^{-1} = (\mathrm{Cov}(\nu_t)_{i,i})^{-1}(1 - s_i \mathbf{b}(\nu_t)_i)$, we finally get

$$\mathbb{E}[\mathbf{b}(\nu_{t+h}) - \mathbf{b}(\nu_t)|u(t)] = h \mathbf{\Psi}(\nu_t) P_t Q_t v + o(h),$$

where $S = \mathrm{diag}(\mathrm{Sgn}(v))$, $P_t = \mathbf{I}_n - S\mathrm{diag}(\mathbf{b}(\nu_t))$ and $Q_t = \mathbf{I}_n + S\mathrm{diag}(\mathbf{b}(\mu_t))$. Integrating with respect to time gives

$$\mathbf{b}(\mu) - \mathbf{b}(\nu) \overset{(65)}{=} \mathbb{E}\mathbf{b}(\nu_1) - \mathbf{b}(\nu_0) = \int_0^1 \mathbb{E}\left[\mathbf{\Psi}(\nu_t) P_t Q_t v\right] dt. \tag{66}$$

Since $\nu_t$ is a pinning of $\nu$, the assumption (51) implies that $\|\mathbf{Cor}(\nu_t)\|_{\mathrm{OP}} \leq \alpha$. The triangle inequality then gives

$$|\mathbf{b}(\mu) - \mathbf{b}(\nu)| \leq \alpha \int_0^1 \mathbb{E}|\left(\mathbf{I}_n - S\mathrm{diag}(\mathbf{b}(\nu_t))\right)\left(\mathbf{I}_n + S\mathrm{diag}(\mathbf{b}(\mu_t))\right) v| dt. \tag{67}$$

Since $|\left(\mathbf{I}_n - S\mathrm{diag}(\mathbf{b}(\nu_t))\right)\left(\mathbf{I}_n + S\mathrm{diag}(\mathbf{b}(\mu_t))\right) v| \leq 4|v|$, this proves the first part of the lemma.

For the second part of the lemma, we write

$$\langle v, \mathbf{b}(\mu) - \mathbf{b}(\nu)\rangle \overset{(66)}{=} \int_0^1 \mathbb{E}\left[\langle v, \mathbf{\Psi}(\nu_t) P_t Q_t v\rangle\right] dt.$$

Define $E_t = \operatorname{diag}(\operatorname{Cov}(\nu_t))$. We have

$$\begin{aligned}(E_t)_{ii} &= 1 - \mathbf{b}_i(\nu_t)^2 \\ &= (1 - \mathbf{b}_i(\nu_t))(1 + \mathbf{b}_i(\nu_t)) \le 2(1 + \mathbf{b}_i(\nu_t)).\end{aligned} \tag{68}$$

Recall that by definition $\mathbf{\Psi}(\nu_t) = \operatorname{Cov}(\nu_t) E_t^{-1}$ and $\mathbf{Cor}(\nu_t) = E_t^{-1/2} \operatorname{Cov}(\nu_t) E_t^{-1/2}$. We have

$$\begin{aligned}\langle v, \mathbf{\Psi}(\nu_t) P_t Q_t v\rangle &= \left\langle E_t^{1/2} v, \mathbf{Cor}(\nu_t) E_t^{-1/2} P_t Q_t v \right\rangle \\ &\le \|\mathbf{Cor}(\nu_t)\|_{\mathrm{OP}} \cdot \left|E_t^{1/2} v\right| \cdot \left|E_t^{-1/2} P_t Q_t v\right|.\end{aligned} \tag{69}$$

The following diagonal matrix is entry-wise bounded.

$$\begin{aligned}(E_t^{-1/2} P_t Q_t)_{ii} &= \frac{(1 - s_i \mathbf{b}_i(\nu_t))\,(1 + s_i \mathbf{b}_i(\nu_t))}{(1 - \mathbf{b}_i^2(\nu_t))^{1/2}} \\ &\le 2K \exp\left(2\,|v_i|\right) \sqrt{\frac{1 + \mathbf{b}_i(\nu_t)}{1 - \mathbf{b}_i(\nu_t)}} \\ &\le 2K^{3/2} C^{1/2} \exp\left(2\,|v_i|\right) (1 + \mathbf{b}_i(\nu))^{1/2}.\end{aligned}$$

The last step follows because: if $i \notin F_t$, meaning that $i$-th coordinate has been pinned, by the convention on $\frac{0}{0}$, $(E_t^{-1/2} P_t Q_t)_{ii} \le 1$; if $s_i = -1$, then we bound $(1 + s_i \mathbf{b}_i(\mu_t))$ from above by 2; otherwise, applying Lemma 72 to $\nu_t$ together with condition (52), we obtain

$$1 + \mathbf{b}(\mu_t)_i = 1 + \mathbf{b}_i(\mathcal{T}_{(1-t)v} \nu_t) \le 2K \frac{1 + \mathbf{b}_i(\nu_t)}{1 - \mathbf{b}_i(\nu_t)} \exp\left(2\,|v_i|\right).$$

Plugging the above bound together with (68) back to (69), we obtain

$$\begin{aligned}&\langle v, \mathbf{\Psi}(\nu_t) P_t Q_t v\rangle \\ &\le \|\mathbf{Cor}(\nu_t)\|_{\mathrm{OP}} \sqrt{\sum_i 2(1 + \mathbf{b}_i(\nu_t)) v_i^2} \sqrt{\sum_i 4K^3 C \exp(4\,|v_i|)\,(1 + \mathbf{b}_i(\nu))\, v_i^2} \\ &\le \|\mathbf{Cor}(\nu_t)\|_{\mathrm{OP}} \left(\sum_i 4K^3 C \exp(4\,|v_i|)\,(1 + \mathbf{b}_i(\nu))\, v_i^2\right)\end{aligned}$$

where the last step follows from condition (52). Hence,

$$\langle v, \mathbf{b}(\mu) - \mathbf{b}(\nu)\rangle \le 4\alpha \sum_{i \in [n]} K^3 C\,(1 + \mathbf{b}_i(\nu))\, v_i^2 \exp(4|v_i|),$$

which completes the proof. □

*Proof of lemma 69.* Define $f(v) := \log \int_{\{-1,1\}^n} \exp(\langle v, x\rangle) d\nu(x)$ for $v \in \mathbb{R}^n$, the logarithmic Laplace transform of $\nu$. Define for $x \in \mathbb{R}^n$

$$g(x) = \max_{v \in \mathbb{R}^n} \langle v, x\rangle - f(v),$$

its Legendre dual. According to Lemma 33, we have $\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_v \nu || \nu) = g(\nabla f(v))$ and that $\mathbf{b}(\mathcal{T}_v \nu) = \nabla f(v)$. With this notation, the assumption (54) amounts to

$$|\nabla f(v') - \nabla f(0)|^2 \leq \varepsilon^2 |v'|^2, \quad \forall v' \in \mathbb{R}^n, \tag{70}$$

and the conclusion becomes

$$g(\nabla f(v)) \geq \frac{1}{2\varepsilon} |\nabla f(v) - \nabla f(0)|^2, \forall v \in \mathbb{R}^n. \tag{71}$$

Define $h(v) = f(v) - \langle \nabla f(0), v\rangle$. Equation (70) implies

$$h(v) = \int_0^1 \langle \nabla h(tv), v\rangle dt \leq |v| \int_0^1 |\nabla h(tv)| dt \leq |v| \int_0^1 \varepsilon t |v| dt = \frac{1}{2}\varepsilon |v|^2.$$

Since the Legendre transform is order-reversing, we have

$$\sup_{v \in \mathbb{R}^n} \langle x, v\rangle - h(v) \geq \frac{1}{2\varepsilon}|x|^2, \quad \forall x \in \mathbb{R}^n$$

which yields (71) by taking $x = \nabla f(v) - \nabla f(0)$ and completes the proof for the first part of the lemma.

For the second part, note that the assumption (55) amounts to

$$\langle v, \nabla f(v) - \nabla f(0)\rangle \leq \sum_{i \in [n]} \varepsilon_i v_i^2 \exp(4|v_i|), \quad \forall v \in \mathbb{R}^n.$$

Define $h(v) = f(v) - \langle \nabla f(0), v\rangle$, then

$$\langle v, \nabla h(v)\rangle \leq \sum_{i \in [n]} \varepsilon_i v_i^2 \exp(4|v_i|), \quad \forall v \in \mathbb{R}^n.$$

We can upper bound $h$ as follows

$$\begin{aligned}
h(v) &= h(0) + \int_0^1 \langle v, \nabla h(tv)\rangle \, dt \\
&\leq \int_0^1 \sum_{i \in [n]} \varepsilon_i v_i^2 t \exp(4t|v_i|) dt \\
&= \sum_{i \in [n]} \frac{1}{16} \varepsilon_i \left[\exp(4|v_i|)(4|v_i| - 1) + 1\right] \\
&\leq \frac{1}{16} \sum_{i \in [n]} \varepsilon_i \left[\exp(8|v_i|) - 8|v_i| - 1\right] \\
&\leq \frac{1}{16} \sum_{i \in [n]} \varepsilon_i \left[\exp(16|v_i|) - 16|v_i| - 1\right]
\end{aligned}$$

The Legendre transform of the function $s \mapsto \frac{\varepsilon}{16}\left(\exp(|16s|) - |16s| - 1\right)$ is the function $t \mapsto \frac{\varepsilon}{16}\left[\left(1+\frac{|t|}{\varepsilon}\right)\log\left(1+\frac{|t|}{\varepsilon}\right) - \frac{|t|}{\varepsilon}\right]$. Since the Legendre transform is order-reversing, we get

$$h^*(x) \geq \sum_{i\in[n]} \frac{\varepsilon_i}{16}\left(\left(1+\frac{|x_i|}{\varepsilon_i}\right)\log\left(1+\frac{|x_i|}{\varepsilon_i}\right) - \frac{|x_i|}{\varepsilon_i}\right).$$

Let $y = \nabla f(0)$. Taking $\varepsilon = \frac{\varepsilon_i}{1+y_i} \geq 1$ in equation (58) of Lemma 71, we obtain

$$\frac{1+y_i}{4\varepsilon_i} H(x_i+y_i, y_i) \overset{(58)}{\leq} \varepsilon_i \Phi\left(\frac{x_i}{\varepsilon_i}\right) \overset{(59)}{\leq} 3\varepsilon_i \Phi\left(\frac{|x_i|}{\varepsilon_i}\right).$$

Finally, using the assumption $\varepsilon_i \leq C\left(1+\mathbf{b}_i(\nu)\right)$ and taking $x = \nabla f(v) - \nabla f(0)$, we get

$$\sum_{i\in[n]} H(x_i+y_i, y_i) \leq 192C \cdot h^*(x) = 192C \cdot g(x+y) = 192C \cdot \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_v\nu||\nu).$$

This completes the proof. □

# References


[AASV21] Yeganeh Alimohammadi, Nima Anari, Kirankumar Shiragur, and Thuy-Duong Vuong. Fractionally log-concave and sector-stable polynomials: counting planar matchings and more. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 433–446, 2021.

[AJK+21a] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence ii: optimal sampling and concentration via restricted modified log-Sobolev inequalities. *arXiv preprint arXiv:2111.03247*, 2021.

[AJK+21b] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence in high-dimensional expanders: modified log-Sobolev inequalities for fractionally log-concave polynomials and the Ising model. *arXiv preprint arXiv:2106.04105*, 2021.

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *STOC '20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1198–1211. ACM, New York, [2020] ©2020.

[ALGV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12. ACM, New York, 2019.

[ALOG20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pages 1319–1330. IEEE Computer Soc., Los Alamitos, CA, 2020.

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

[AMS22] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. *arXiv preprint arXiv:2203.05093*, 2022.

[BB19] Roland Bauerschmidt and Thierry Bodineau. A very simple proof of the LSI for high temperature spin systems. *Journal of Functional Analysis*, 276(8):2582–2588, 2019.

[BCC+22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: coupling, spectral independence, and entropy factorization. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692. SIAM, 2022.

[BD22] Roland Bauerschmidt and Benoit Dagallier. Log-Sobolev inequality for near critical Ising models. *arXiv preprint arXiv:2202.02301*, 2022.

[BE19] Sébastien Bubeck and Ronen Eldan. The entropic barrier: exponential families, log-concave geometry, and self-concordance. *Mathematics of Operations Research*, 44(1):264–276, 2019.

[CFYZ21a] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing time for the Ising model in the uniqueness regime. *arXiv preprint arXiv:2111.03034*, 2021.

[CFYZ21b] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of Glauber dynamics via spectral independence for all degrees. *arXiv preprint arXiv:2105.15005*, 2021.

[CFYZ22] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. *arXiv preprint arXiv:2203.07771*, 2022.

[Che21] Yuansi Chen. An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geometric and Functional Analysis*, 31(1):34–61, 2021.

[CLV20] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, pages 1307–1318. IEEE Computer Soc., Los Alamitos, CA, 2020.

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1537–1550, 2021.

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pages 974–985. IEEE Computer Soc., Los Alamitos, CA, 2017.

[DSS22] Jian Ding, Jian Song, and Rongfeng Sun. A new correlation inequality for Ising models with external fields. *Probability Theory and Related Fields*, pages 1–16, 2022.

[EHMT17] Matthias Erbar, Christopher Henderson, Georg Menz, and Prasad Tetali. Ricci curvature bounds for weakly interacting Markov chains. *Electronic Journal of Probability*, 22:1–23, 2017.

[EK11] Ronen Eldan and Bo'az Klartag. Approximately Gaussian marginals and the hyperplane conjecture. In *Concentration, functional inequalities and isoperimetry*, volume 545 of *Contemp. Math.*, pages 55–68. Amer. Math. Soc., Providence, RI, 2011.

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

[Eld] Ronen Eldan. Lecture notes – from stochastic calculus to geometric inequalities. https://www.wisdom.weizmann.ac.il/ ronene/GFANotes.pdf.

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

[Eld20] Ronen Eldan. Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation. *Probability Theory Related Fields*, 176(3-4):737–755, 2020.

[Eld22] Ronen Eldan. Analysis of high-dimensional distributions using pathwise methods. In *Proceedings of ICM, to appear*, 2022.

[EMZ20] Ronen Eldan, Dan Mikulincer, and Alex Zhai. The CLT in high dimensions: quantitative bounds via martingale embedding. *Annals of Probability*, 48(5):2494–2524, 2020.

[ES22] Ronen Eldan and Omer Shamir. Log concavity and concentration of Lipschitz functions on the Boolean hypercube. *Journal of Functional Analysis*, 282(8):Paper No. 109392, 2022.

[FGYZ21] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the Boolean domain. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1558–1577. SIAM, 2021.

[GL10] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.

[GM87] M. Gromov and V. D. Milman. Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces. *Compositio Math.*, 62(3):263–282, 1987.

[JSTV04] Mark Jerrum, Jung-Bae Son, Prasad Tetali, and Eric Vigoda. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.*, 14(4):1741–1765, 2004.

[Kla06] B. Klartag. On convex perturbations with a bounded isotropic constant. *Geom. Funct. Anal.*, 16(6):1274–1290, 2006.

[Kla17] Bo'az Klartag. Needle decompositions in Riemannian geometry. *Mem. Amer. Math. Soc.*, 249(1180):v+77, 2017.

[Kla18] Bo'az Klartag. Eldan's stochastic localization and tubular neighborhoods of complex-analytic sets. *Journal of Geometric Analysis*, 28(3):2008–2027, 2018.

[KLS95a] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete & Computational Geometry*, 13(3-4):541–559, 1995.

[KLS95b] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.*, 13(3-4):541–559, 1995.

[KM12] Bo'az Klartag and Emanuel Milman. Centroid bodies and the logarithmic Laplace transform—a unified approach. *J. Funct. Anal.*, 262(1):10–34, 2012.

[KM17] Tali Kaufman and David Mass. High dimensional random walks and colorful expansion. In *8th Innovations in Theoretical Computer Science Conference*, volume 67 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 4, 27. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[KM21] B Klartag and V Milman. The slicing problem by Bourgain. *To appear in Analysis at Large, A Collection of Articles in Memory of Jean Bourgain*, 2021.

[KO20] Tali Kaufman and Izhar Oppenheim. High order random walks: beyond spectral gap. *Combinatorica*, 40(2):245–281, 2020.

[Liu21] Kuikui Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. *arXiv preprint arXiv:2103.11609*, 2021.

[Lov99] László Lovász. Hit-and-run mixes fast. *Mathematical Programming*, 86(3, Ser. A):443–461, 1999.

[LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition, with contributions by Elizabeth L. Wilmer, with a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[LST21] Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Structured logconcave sampling with a restricted Gaussian oracle. In *Conference on Learning Theory*, pages 2993–3050. PMLR, 2021.

[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)*, pages 998–1007. IEEE, 2017.

[Mil21] Emanuel Milman. The quasi curvature-dimension condition with applications to sub-Riemannian manifolds. *Comm. Pure Appl. Math.*, 74(12):2628–2674, 2021.

[MR02] Neal Madras and Dana Randall. Markov chain decomposition for convergence rate analysis. *The Annals of Applied Probability*, 12(2):581 – 606, 2002.

[NS22] Hariharan Narayanan and Piyush Srivastava. On the mixing time of coordinate hit-and-run. *Combinatorics, Probability and Computing*, 31(2):320–332, 2022.

[Opp18] Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders Part I: descent of spectral gaps. *Discrete & Computational Geometry*, 59(2):293–330, 2018.

[STL20] Ruoqi Shen, Kevin Tian, and Yin Tat Lee. Composite logconcave sampling with a restricted Gaussian oracle. *arXiv preprint arXiv:2006.05976*, 2020.

# A Appendix A: Loose ends

*Proof of Lemma 36.* Let $K$ and $v(x)$ be defined as in Lemma 33. Define $z = \mathrm{b}(\nu)$. A direct calculation shows that

$$H(z,z) = 0, \quad \nabla_x H(x,z)|_{x=z} = 0.$$

Moreover, we have

$$\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{v(z)}\nu||\nu) = 0, \quad \nabla_x \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{v(x)}\nu||\nu) = 0|_{x=z} = 0.$$

Therefore, if we establish that

$$\nabla_x^2 H(x,z) \preceq \alpha \nabla_x^2 \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{v(x)}\nu||\nu), \ \ \forall x \in K \tag{72}$$

Then it will follow that $H(x,z) \leq \alpha \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{v(x)}\nu||\nu)$ which will complete the proof. A direct calculation yields

$$\frac{\partial^2}{\partial x_i^2} H(x,z) = \frac{1}{1-x_i^2}.$$

Since $\nu$ is supported on $\{-1,1\}^n$, we have $\int x^{\otimes 2} d\nu(x) = \mathbf{I}_n$, which implies that

$$\mathrm{diag}(\mathrm{Cov}(\mathcal{T}_{v(x)}\nu)) = \mathbf{I}_n - x^{\otimes 2}.$$

A combination of the last two displays gives

$$\nabla_x^2 H(x,z) = \mathrm{diag}(\mathrm{Cov}(\mathcal{T}_{v(x)}\nu))^{-1}.$$

Moreover, Formula (22) gives

$$\nabla_x^2 \mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{v(x)}\nu||\nu) = \mathrm{Cov}(\mathcal{T}_{v(x)}\nu)^{-1}. \tag{73}$$

Combining the last two displays with the assumption $\mathbf{Cor}(\mathcal{T}_{v(x)}) \preceq \alpha \mathbf{I}_n$ implies (72). The proof is complete. □

*Proof of Fact 23.* Let $X \sim \nu$. We have

$$\begin{aligned}
2\mathrm{Cov}(\nu)_{i,j} &= \mathbb{E}[X_i|X_j = 1](1+\mathbb{E}[X_j]) - \mathbb{E}[X_i|X_j=-1](1-\mathbb{E}[X_j]) - 2\mathbb{E}[X_i]\mathbb{E}[X_j] \\
&= (\mathbb{E}[X_i|X_j=1] - \mathbb{E}[X_i|X_j=-1]) + \mathbb{E}[X_j]\Big(\mathbb{E}[X_i|X_j=1] + \mathbb{E}[X_i|X_j=-1] \\
&\quad -\mathbb{E}[X_i|X_j=1](1+\mathbb{E}[X_j]) - \mathbb{E}[X_i|X_j=-1](1-\mathbb{E}[X_j])\Big) \\
&= (\mathbb{E}[X_i|X_j=1] - \mathbb{E}[X_i|X_j=-1])(1-\mathbb{E}[X_j]^2),
\end{aligned}$$

which readily implies that $\mathbf{\Psi}(\nu) = \mathrm{Cov}(\nu)D^{-1}$ where $D := \mathrm{diag}(\mathrm{Cov}(\nu))$. To see that the operator norm is that same as that of $\mathbf{Cor}(\nu) = D^{-1/2}\mathrm{Cov}(\nu)D^{-1/2}$, let $v$ be an eigenvector of $\mathbf{Cor}(\nu)$ with eigenvalue $\lambda$, write $u = D^{1/2}v$, then

$$\lambda v = D^{-1/2}\mathrm{Cov}(\nu)D^{-1/2}v \Leftrightarrow \lambda D^{-1/2}u = D^{-1/2}\mathrm{Cov}(\nu)D^{-1}u \Leftrightarrow \lambda u = \mathrm{Cov}(\nu)D^{-1}u$$

which implies that $\rho(\mathbf{\Psi}(\nu)) = \rho(\mathbf{Cor}(\nu))$. Since $\mathbf{Cor}(\nu)$ is symmetric, we have $\rho(\mathbf{Cor}(\nu)) = \|\mathbf{Cor}(\nu)\|_{\mathrm{OP}}$. This completes the proof. □

## A.1 Supermartingality of the Dirichlet form: Proof of Proposition 48

Recall that we assume that $\mathcal{L}_f$ is a Doob localization. Let $P$ be the transition kernel of the Markov chain given by $(\mathcal{L}_f, t, \mu)$, then there exists a random variable $X \sim \mu$ and a $\sigma$-algebra $\Sigma$ on $X$ such that

$$\begin{aligned}\int_{\Omega\times\Omega} \varphi(x)\varphi(y)dP_x(y)d\mu(x) &= \mathbb{E}\left[\int_{\Omega\times\Omega} \varphi(x)\varphi(y)\mu_t(dx)\mu_t(dy)\right] \\ &= \mathbb{E}_\Sigma\left[\mathbb{E}_X\left[\mathbb{E}[\varphi(X)|\Sigma]^2\right]\right].\end{aligned}$$

Moreover,

$$\int_{\Omega\times\Omega} Pf(x)\log Pf(x)d\mu(x) = \mathbb{E}_\Sigma\left[\mathbb{E}_X\left[\mathbb{E}[f(X)|\Sigma]\log\mathbb{E}[f(X)|\Sigma]\right]\right] - \mathbb{E}_\nu[f]\log\mathbb{E}_\nu[f].$$

Fix a measure $\rho$ on $\Omega$ and suppose that $\mu$ is absolutely continuous with respect to $\rho$, hence we may write $\frac{d\mu}{d\rho}(x) = h(x)$.

**Lemma 73.** *Fix a measure $\rho$ on $\Omega$ and $\varphi : \Omega \to \mathbb{R}$ and a $\sigma$-algebra $\Sigma$ on $\Omega$. For all $x \in \Omega$ there exist linear functionals $F_x, G_x : L_1(\Omega,\rho) \to \mathbb{R}$ such that the following holds: For every $h : \Omega \to \mathbb{R}_+$ satisfying $\int_\Omega h d\rho = 1$,*

$$\left[\mathbb{E}_X\left[\mathbb{E}[\varphi(X)|\Sigma]^2\right]\right] = \int_\Omega \frac{F_x(h)^2}{G_x(h)}d\rho(x).$$

*Analogously, for every non-negative $f : \Omega \to \mathbb{R}$, there exist families of linear functionals $F_x, G_x$ such that*

$$\left[\mathbb{E}_X\left[\mathbb{E}[f(X)|\Sigma]\log\mathbb{E}[f(X)|\Sigma]\right]\right] = \int_\Omega F_x(h)\log\left(F_x(h)/G_x(h)\right)$$

*where $X \sim \mu$ and $\frac{d\mu}{d\rho} = h$.*

*Proof.* The lemma follows from a standard disintegration theorem. For the sake of demystifying the proof, let us first consider the case that $\Sigma$ is a finite partition of $\Omega$ into sets $\Omega_1, ..., \Omega_k$. Let $A(x)$ be the unique set $\Omega_i$ such that $x \in \Omega_i$. Then we have

$$\mathbb{E}[\varphi(X)|\Sigma] = \frac{\int_{A(X)}\varphi(x)d\mu(x)}{\mu(A(X))}.$$

Therefore,

$$\begin{aligned}\mathbb{E}_{X\sim\mu}\left[\mathbb{E}[\varphi(X)|\Sigma]^2\right] &= \sum_{i=1}^{k}\mu(\Omega_i)\mathbb{E}_{X\sim\mu}\left[\frac{\int_{A(X)}\varphi(x)d\mu(x)}{\mu(A(X))}\middle| X\in\Omega_i\right]^2\\ &= \sum_{i=1}^{k}\frac{\left(\int_{\Omega_i}\varphi(x)d\mu(x)\right)^2}{\mu(\Omega_i)}\\ &\sum_{i=1}^{k}\frac{\left(\int_{\Omega_i}\varphi(x)h(x)d\rho(x)\right)^2}{\int_{\Omega_i}h(x)\rho(dx)}.\end{aligned}$$

This completes the proof of the first part by choosing

$$F_x(h) := \frac{1}{\rho(A(x))}\int_{A(x)}\varphi(y)h(y)d\rho(y) \quad \text{and} \quad G_x(h) := \frac{1}{\rho(A(x))}\int_{A(x)}h(y)d\rho(y).$$

The proof of the second part is very similar. Define $\psi(s) = s\log s$. We have,

$$\begin{aligned}\mathbb{E}_{X\sim\mu}\left[\mathbb{E}[f(X)|\Sigma]\log\mathbb{E}[f(X)|\Sigma]\right] &= \sum_{i=1}^{k}\mu(\Omega_i)\psi\left(\mathbb{E}_{X\sim\mu}\left[\frac{\int_{A(X)}f(x)d\mu(x)}{\mu(A(X))}\middle| X\in\Omega_i\right]\right)\\ &= \sum_{i=1}^{k}\nu(\Omega_i)\psi\left(\frac{\left(\int_{\Omega_i}f(x)d\mu(x)\right)}{\mu(\Omega_i)}\right)\\ &\sum_{i=1}^{k}\left(\int_{\Omega_i}f(x)h(x)d\rho(x)\log\left(\frac{\int_{\Omega_i}f(x)h(x)\rho(dx)}{\int_{\Omega_i}h(x)d\rho(x)}\right)\right),\end{aligned}$$

and the second part is proved by choosing

$$F_x(h) := \frac{1}{\rho(A(x))}\int_{A(x)}f(y)h(y)d\rho(y) \quad \text{and} \quad G_x(h) := \frac{1}{\rho(A(x))}\int_{A(x)}h(y)d\rho(y).$$

□

To complete the proof, observe that for every linear functional $F : L_1(\Omega,\nu)\to\mathbb{R}$, we have that $t\to F(\frac{d\nu_t}{d\nu})$ is a martingale.

**Fact 74.** *If $M_t, N_t$ are martingales and $N_t\geq 0$ almost surely, then $\frac{N_t^2}{M_t}$ is a submartingale. If $M_t$ is also positive almost-surely then $N_t\log\frac{N_t}{M_t}$ is a submartingale.*

*Proof.* The first part follows immediately from the fact that the function $(x,y)\to x^2/y$ is convex in the domain $\{y>0\}$. The second part follows from the convexity of $(x,y)\to x\log(x/y)$ in the domain $\{x>0, y>0\}$. □

## A.2 Existence and entropic decay of the negative-fields localization process

*Proof of Proposition 18.* For $v \in \mathbb{R}^n$ and $u \in \{-1, 0, 1\}^n$, define

$$a(v, u) = \mathbf{b}(\mathcal{R}_u \mathcal{T}_v \nu).$$

Fix $t \geq 0$ and suppose that the process $u(t)$ has been constructed up to time $t$. We will first show that the process can be extended up to some stopping time $\tau > t$. Define

$$F(t) := \{i \in [n]; u_t(i) = 0\},$$

understood as the set of coordinates that have not been pinned. For any $i \in F_t$ consider a random variable $T_i$ defined by the formula

$$\mathbb{P}(T_i > s) = \exp\left(-\int_t^s (1 - s_i(r) a_i(v(r), u(t)))|v_i'(r)|dr\right),$$

where $s_i(r) := \operatorname{sign}(v_i'(r))$ (note that the right-hand side is $u(t)$-measurable), and such that the $T_i$'s are independent. Define

$$\tau := \min_{i \in F_t} T_i \quad \text{and} \quad J := \arg\min_{i \in F_t} T_i.$$

We can now extend the process up to time $\tau$ by setting

$$u(s) = \begin{cases} u(t), & t \leq s < \tau \\ u(t) - s_J(\tau)\mathbf{e}_J, & s = \tau, \end{cases}$$

where $\mathbf{e}_J$ is the $J$-th standard basis vector. In words, up to the stopping time $\tau$, we apply an exponential tilt to the measure according to $v(\cdot)$, and at time $\tau$ we pin the $j$-th coordinate according to the sign of $v_j'(\tau)$. We can now define the process for all times by iteratively extending it until the next stopping time. Since at any such stopping time, one of the coordinates is pinned, we only need to repeat this iteration at most $n$ times and thus the process is well-defined for $t \in [0, \infty)$.

Next, we show that the process is a martingale. First observe that the process is time-equivariant in the sense that the evolution of $(\nu_s)_{s \geq t}$ conditioned on $(\nu_r)_{r \in [0,s]}$ is the same as the evolution of $(\tilde{\nu}_t)_t$, the process obtained from the starting measure $\tilde{\nu} = \nu_s$ and using $\tilde{v}(t) = v(t + s) - v(s)$ as the driving curve. Indeed, this follows from the memoryless property of exponential random variables, which implies that for every $t > 0$ and every $i \in F_t$, we have that $T_i | T_i > s$ has the same distribution as $\tilde{T}_i$.

In order to show that the process is a martingale, we need to show that for every $s, t \geq 0$ we have $\mathbb{E}[\nu_{t+s} | \nu_s] = \nu_s$ and by the above discussion it is enough to consider the case $s = 0$, and show that $\mathbb{E}[\nu_t(x)] = \nu(x)$. We claim that it is yet enough to show that

$$\mathbb{E}[\nu_h(x)] = \nu(x) + o(h), \quad \forall h \geq 0. \tag{74}$$

Indeed, if this is the case then we can write

$$\mathbb{E}[\nu_t] = \sum_{i=1}^{k} \mathbb{E}\left[\mathbb{E}\left[\nu_{it/k} - \nu_{(i-1)t/k} | \nu_{(i-1)t/k}\right]\right] \leq k o(1/k),$$

for all $k \in \mathbb{N}$, which implies that $\mathbb{E}[\nu_t(x)] = \nu(x)$.

Let us now prove (74). By reflecting both $\nu$ and $v(\cdot)$ around coordinate directions, we may clearly assume without loss of generality that $v_i'(0) \geq 0$ for all $i \in [n]$. Since for all $i$ we have $\mathbb{P}(T_i) \in [0,h] = O(h)$, we can write

$$\begin{aligned}
\nu_h &= \mathcal{T}_{v(h)} \left( \mathbf{1}_{\{\tau > h\}} \nu + \mathbf{1}_{\{\tau \in [0,h]\}} \mathcal{R}_{-\operatorname{sign}(v_j'(\tau))\mathbf{e}_J} \nu \right) + o(h) \\
&= \mathcal{T}_{hv'(0)} \left( \mathbf{1}_{\{\tau > h\}} \nu + \sum_{i \in [n]} \mathbf{1}_{\{T_i \in [0,h]\}} \mathcal{R}_{-\mathbf{e}_i} \nu \right) + o(h) \\
&= \mathcal{T}_{hv'(0)} \left( \nu + \sum_{i \in [n]} \mathbf{1}_{\{T_i \in [0,h]\}} \left( \mathcal{R}_{-\mathbf{e}_i} \nu - \nu \right) \right) + o(h).
\end{aligned}$$

Now, note that for all $x \in \{-1,1\}^n$,

$$\mathcal{T}_{hv'(0)} \nu(x) = \nu(x)(1 + \langle x - \mathbf{b}(\nu), hv'(0) \rangle) + o(h),$$

and

$$\mathcal{R}_{-\mathbf{e}_i} \nu(x) = \nu(x) \left( 1 - \frac{x_i - \mathbf{b}(\nu)_i}{1 - \mathbf{b}(\nu)_i} \right).$$

Combining the last three displays gives

$$\begin{aligned}
\mathbb{E}[\nu_h(x)] &= \nu(x) \left( 1 + \langle x - \mathbf{b}(\nu), hv'(0) \rangle - \sum_{i \in [n]} \mathbf{1}_{\{T_i \in [0,h]\}} \frac{x_i - \mathbf{b}(\nu)_i}{1 - \mathbf{b}(\nu)_i} \right) + o(h) \\
&= \nu(x) \left( 1 + \sum_{i \in [n]} (x_i - \mathbf{b}(\nu)_i) \left( hv_i'(0) - \frac{\mathbf{1}_{\{T_i \in [0,h]\}}}{1 - \mathbf{b}(\nu)_i} \right) \right) + o(h), \qquad (75)
\end{aligned}$$

at which point we have established the correctness of (8). By definition of the random variables $T_i$, we have

$$\mathbb{P}(T_i \in [0,h]) = h(1 - \mathbf{b}(\nu)_i) v_i'(0) + o(h),$$

and therefore

$$\mathbb{E}\left[ hv_i'(0) - \frac{\mathbf{1}_{\{T_i \in [0,h]\}}}{1 - \mathbf{b}(\nu)_i} \right] = o(h).$$

Plugging the last display into (75) implies that $\mathbb{E}[\nu_h(x)] = \nu(x) + o(h)$ which proves (74), so we have established that the process is a martingale.

Under the extra condition that $\lim_{t \to \infty} |v_i(t)| = \infty$, we just observe that $\nu_t$ converges to a Dirac measure almost-surely under every pinning process $u(t)$, and therefore the process is a localization process. □

*Proof of Proposition 41.* Fix a measure $\nu$ on $\{-1,1\}^n$ and let $(\nu_t)_t$ be the process obtained via the negative-fields localization process. Fix $t > 0$. According to equation (7), we can write

$$\nu_{t+h}(x) = \nu_t(x) \left( 1 + \langle x - \mathbf{b}(\nu_t), Z \rangle \right) + o(h), \qquad (76)$$

where, conditional on $\nu_t$, the random variable $Z$ has independent coordinates which satisfy

$$Z_i = \begin{cases} -h & \text{with probability } 1 - h(1 + \mathbf{b}(\nu_t)_i) \\ \frac{1}{1 + b(\nu_t)_i} & \text{with probability } h(1 + \mathbf{b}(\nu_t)_i). \end{cases}$$

Let $f : \{-1,1\}^n \to \mathbb{R}_+$ be such that $\int f d\nu > 0$. We have,

$$\begin{aligned}\int f(x) d\nu_{t+h}(x) &= \int f(x) d\nu_t(x) \left(1 + \left\langle \left( \frac{\int x f(x) d\nu_t(x)}{\int x f(x) \nu_t(dx)} - \mathbf{b}(\nu_t) \right), Z \right\rangle \right) + o(h) \\ &= \nu_t(f) \left(1 + \langle V, Z \rangle\right) + o(h),\end{aligned}$$

where

$$v := \frac{\int x f(x) \nu_t(dx)}{\int f(x) \nu_t(dx)} - \mathbf{b}(\nu_t) = \mathbf{b}(\mu_t) - \mathbf{b}(\nu_t)$$

and where

$$\frac{d\mu_t}{d\nu_t}(x) := \frac{f(x)}{\int f(x) d\nu_t(x)}.$$

Therefore, we have

$$\mathbb{E}\left[\nu_{t+h}(f) \log \nu_{t+h}(f)\right] - \nu_t(f) \log \nu_t(f) = \nu_t(f) \mathbb{E}\left[(1 + \langle Z, v \rangle) \log(1 + \langle Z, v \rangle)\right] + o(h). \quad (77)$$

For a fixed $t$, define $b_i = \mathbf{b}(\nu_t)_i$. We calculate,

$$\begin{aligned}&\mathbb{E}\left[(1 + \langle Z, v \rangle) \log(1 + \langle Z, v \rangle)\right] \\ &= -h \sum_{i=1}^n v_i + h \sum_{i=1}^n (1 + b_i) \left(1 + \frac{v_i}{1 + b_i}\right) \log\left(1 + \frac{v_i}{1 + b_i}\right) + o(h) \\ &= h \sum_{i=1}^n (1 + b_i) \left( \left(1 + \frac{v_i}{1 + b_i}\right) \log\left(1 + \frac{v_i}{1 + b_i}\right) - \frac{v_i}{1 + b_i} \right) + o(h) \\ &\overset{(57)}{\leq} 4h H(b + v, b) + o(h),\end{aligned}$$

Combining the two last displays yields

$$\mathbb{E}[\mathrm{Ent}_{\nu_{t+h}}[f] | \nu_t] \geq \mathrm{Ent}_{\nu_t}[f] - 4h \nu_t(f) H(b + v, b) + o(h)$$

The assumption that $\nu$ is $\alpha$-entropically stable with respect to $H(x, y)$ together with Lemma 31 give

$$H(b + v, b) \leq \alpha \mathrm{D}_{\mathrm{KL}}(\mu_t || \nu_t) = \alpha \frac{\mathrm{Ent}_{\nu_t}[f]}{\nu_t(f)}.$$

Combining the last two displays completes the proof. □

## A.3 Proofs of technical lemmas from Section 6

*Proof of Lemma 70.* For $k \in [n]$. Let $Z \sim \nu$ and define $X = Z_j$ and $Y = Z_i$. We have

$$\begin{aligned}\mathrm{Cov}(\nu)_{ij} &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X|Y = 1]\mathbb{P}(Y = 1) - \mathbb{E}[X|Y = -1]\mathbb{P}(Y = -1) - \mathbb{E}[X]\mathbb{E}[Y] \\ &= 2\mathbb{E}[X|Y = 1]\mathbb{P}(Y = 1) - \mathbb{E}[X] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \left(\mathbb{E}[X|Y = 1] - \mathbb{E}[X]\right)\left(1 + \mathbb{E}[Y]\right).\end{aligned}$$

Therefore, we have

$$\mathbf{b}\left(\mathcal{R}_{\mathbf{e}_i} \nu\right)_j - \mathbf{b}(\nu)_j = \mathbb{E}[X|Y = 1] - \mathbb{E}[X] = \frac{\mathrm{Cov}(\nu)_{ij}}{1 + \mathbf{b}(\nu)_i}$$

which proves the lemma for the case $s = +1$. The proof for $s = -1$ is similar. □

*Proof of Lemma 71.* We first deal with equation (57). Fixing $y \in (-1, 1)$, we calculate the first derivatives with respect to $x$

$$\frac{\partial}{\partial x} H(x, y) = \frac{1}{2} \log\left(\frac{1+x}{1+y}\right) - \frac{1}{2} \log\left(\frac{1-x}{1-y}\right)$$
$$\frac{\partial}{\partial x}(1+y)\Phi\left(\frac{x-y}{1+y}\right) = \log\left(\frac{1+x}{1+y}\right).$$

The two functions coincide up to the first order Taylor expansion in $x$ at $x = y$. Next we look at the second derivatives

$$\frac{\partial^2}{\partial x^2} H(x, y) = \frac{1}{(1+x)(1-x)}$$
$$\frac{\partial^2}{\partial x^2}(1+y)\Phi\left(\frac{x-y}{1+y}\right) = \frac{1}{1+x}.$$

Since $1 \le 2\frac{1}{1-x}$, we obtain the inequality on the right-hand side. Similarly, since $\frac{1}{1-x} \le 2$ when $x \le 1/2$, we also obtain the inequality on the left-hand side for the case $x \le 1/2$. For $x > 1/2$, let

$$G(x, y) = 2(1+y)\Phi(\frac{x-y}{1+y}) - H(x, y).$$

Its first derivative is as follows

$$\frac{\partial}{\partial x} G(x, y) = \frac{1}{2} \log\left(\frac{(1+x)^3(1-x)}{(1+y)^3(1-y)}\right).$$

We observe that $x \mapsto (1+x)^3(1-x)$ is monotonically decreasing on $[1/2, 1]$. Because $\frac{\partial}{\partial x} G(0, y) < 0$, $x \mapsto G(x, y)$ is either decreasing on $[1/2, 1]$ or increasing then decreasing on $[1/2, 1]$. So it is sufficient to check that $G(\frac{1}{2}, y) \ge 0$ and $G(1, y) \ge 0$ to verify $G(x, y) \ge 0$ for $x \in [1/2, 1]$. We verify that

$$G(1, y) = 3 \log\left(\frac{2}{1+y}\right) - 1 + y \ge 0,$$

because $\frac{\partial}{\partial y} G(1, y) = -3/(1+y) + 1 \le 0$ and $G(1, 0) \ge 0$. And we have

$$G(\frac{1}{2}, y) = \frac{5}{2} \log\left(\frac{1}{1+y}\right) + 2y + \frac{9}{4} \log(3/2) - 1 - \frac{1}{4} \log(1/2) \ge 0,$$

because $\frac{\partial}{\partial y} G(\frac{1}{2}, y) = -\frac{5}{2(1+y)} + 2 \le 0$ and $G(\frac{1}{2}, 0) \ge 0$. This proves (57). To prove (58), we note that by using (57), it is enough to show that

$$2\Phi(\delta s) \ge \delta^2 \Phi(s), \quad \forall \delta \in (0, 1), s \in [-1, \infty),$$

and then take $s = \frac{x-y}{(1+y)} \ge -1$ and $\delta = \varepsilon^{-1}$. To prove the last inequality, let $F(s) = 2\Phi(\delta s) - \delta^2 \Phi(s)$. We have

$$F'(s) = 2\delta\Phi'(\delta s) - \delta^2 \Phi'(s) = 2\delta \log(1 + \delta s) - \delta^2 \log(1 + s).$$
$$F''(s) = 2\delta^2 \frac{1}{1+\delta s} - \delta^2 \frac{1}{1+s}$$

For $s \geq 0$, since $F'(0) = 0$ and $F''(s) \geq 0$, we obtain that $F(s)$ is an increasing function of $s$ and hence $F(s) \geq F(0) = 0$. For $s \in [-1, 0)$, since as a function of $\delta$, $F'(s)$ is decreasing, we have $F'(s) \geq 0$. We conclude that $F(s)$ is an increasing function of $s$ on $[-1, 0)$. Hence, $F(s) \geq F(-1) = 2\Phi(-\delta) - \delta^2$. Since

$$\frac{\partial}{\partial \delta}[2\Phi(-\delta) - \delta^2] = 2(-\delta - \log(1 - \delta)) \geq 0,$$

we have $F(s) \geq 2\Phi(-\delta) - \delta^2 \geq 0$.

Finally, we deal with equation (59). For $s \geq 0$, there is nothing to prove since $\Phi(\cdot) \geq 0$ on $[-1, 1]$. For $s \in [-1, 0)$, let $J(s) = 3\Phi(-s) - \Phi(s)$. It has derivatives

$$J'(s) = -3\log(1 - s) - \log(1 + s)$$
$$J''(s) = \frac{3}{1 - s} - \frac{1}{1 + s}.$$

Since $J''(-1/2) = 0, J''(0) > 0$ and $J''(-1) < 0$, we have that $J'(s)$ is minimized at $1/2$ and that it only crosses 0 once on $(-1, 0)$. To check $J(s) \geq 0$, it is sufficient to check $J(0) \geq 0$ and $J(-1) \geq 0$. We conclude equation (59). □

*Proof of Lemma 72.* Clearly we may assume $u = 0$ (otherwise we just need to prove the lemma on for the restriction of $\nu$ to $S_u$). Denote $v = v_i \mathbf{e}_i + \tilde{v}$ where $\tilde{v} \perp \mathbf{e}_i$. We have

$$\mathcal{T}_v \nu = \mathcal{T}_{v_i \mathbf{e}_i} \mathcal{T}_{\tilde{v}} \nu$$

Define $\mu = \mathcal{T}_{\tilde{v}} \nu$. Note that

$$\mathbf{b}_i(\mu) = \mathbb{E}_{X \sim \mathcal{T}_{\tilde{v}} \nu} \left[ \mathbf{b}_i(\mathcal{R}_{\mathrm{Proj}_{\mathbf{e}_i^\perp} X} \nu) \right].$$

According to equation (60), we have $1 + \mathbf{b}_i(\mathcal{R}_{\mathrm{Proj}_{\mathbf{e}_i^\perp} X} \nu) \leq \delta$. Hence

$$1 + \mathbf{b}_i(\mu) \leq \delta.$$

Define $f(t) = \mathbf{b}_i(\mathcal{T}_{t\mathbf{e}_i} \mu)$. If $v_i < 0$ then $f(v_i) < f(0)$ and there is nothing to prove. Otherwise, we have

$$f(0) \leq -1 + \delta, \quad f'(t) = 1 - f(t)^2.$$

Thus,

$$(1 + f(t))' \leq 2(1 + f(t)).$$

By Gronwall's inequality,

$$1 + f(t) \leq (1 + f(0)) \exp(2t) \leq \delta \exp(2t).$$

Therefore,

$$1 + \mathbf{b}_i(\mathcal{T}_v(\nu)) = 1 + f(v_i) \leq \delta \exp(2v_i).$$

For the second part, let $g(t) = \frac{1+f(t)}{1-f(t)}$. Since $x \mapsto \frac{1+x}{1-x}$ is an increasing function of $x \in (-1,1)$, we have

$$\delta' \le g(0) \le \delta''.$$

If $v_i < 0$, then $f(v_i) < f(0)$ and $g(v_i) < g(0)$. Otherwise, we have

$$g'(t) = \frac{2f'(t)}{(1-f(t))^2} = 2g(t).$$

Integrating the above equation, we obtain

$$g(t) = g(0)\exp(2t) \le \tilde{\delta}\exp(2t).$$

The lower bound is established similarly. □

# B Appendix B: Entropic stability for Ising models under a spectral condition

*Proof of Lemma 50.* Consider the Ising model whose density is

$$\nu(x) \propto \exp\left(\langle Jx, x\rangle + \langle v, x\rangle\right)$$

Since $v$ is arbitrary, it suffices to show that

$$\mathrm{Cov}(\nu) \preceq \frac{1}{1-2\|J\|_{\mathrm{OP}}}. \tag{78}$$

Set $\alpha > 2\|J\|_{\mathrm{OP}}$ whose value will be chosen later. Let $M$ be a matrix such that $M^{-1} + \alpha^{-1}\mathbf{I}_n = (\alpha\mathbf{I}_n - 2J)^{-1}$. By the fact that

$$\mathcal{N}(0, M^{-1}) + \mathcal{N}(0, \alpha^{-1}\mathbf{I}_n) \overset{(d)}{=} \mathcal{N}(0, (\alpha\mathbf{I}_n - 2J)^{-1}),$$

we have

$$\begin{aligned}
\exp\left(-\frac{1}{2}\langle(\alpha\mathbf{I}_n - 2J)x, x\rangle\right) &\propto \int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle y, My\rangle} e^{-\frac{\alpha}{2}|x-y|^2} dy \\
&\propto e^{-\alpha|x|^2/2} \int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle y, (M+\alpha\mathbf{I}_n)y\rangle} e^{\alpha\langle x, y\rangle} dy \\
&\propto e^{-\alpha|x|^2/2} \int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle y, (\alpha^2(2J)^{-1})y\rangle} e^{\alpha\langle x, y\rangle} dy.
\end{aligned}$$

Since $|x|$ is constant on $\{-1,1\}^n$, we get that there is a constant $C(J,v)$ such that

$$\exp\left(\langle(\alpha Jx, x\rangle + \langle v, x\rangle\right) = C(J,v)\int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle y, (\alpha^2(2J)^{-1})y\rangle} e^{\langle x, v+\alpha y\rangle} dy, \quad \forall x \in \{-1,1\}^n.$$

It follows that

$$\begin{aligned}
\nu(x) &\propto \int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle y, (\alpha^2(2J)^{-1})y\rangle} e^{\langle x, v+\alpha\rangle} dy \\
&\propto \int_{\mathbb{R}^n} e^{-\frac{1}{2}\langle y, (\alpha^2(2J)^{-1})y\rangle + \log(Z(v+\alpha y))} \left(Z(v+y\alpha)^{-1} e^{\langle x, v+\alpha y\rangle}\right) dy,
\end{aligned}$$

where $Z(w) = \int_{\{-1,1\}^n} \exp(\langle x, w\rangle)dx$. Since $Z(v+\alpha y)^{-1}e^{\langle x, v+\alpha y\rangle}$ is the density of $\mathcal{T}_{v+\alpha y}\mu$ where $\mu$ is the uniform measure, we have

$$\nu = \int_{\mathbb{R}^n} f(y)\mathcal{T}_{v+\alpha y}\mu dy,$$

where

$$f(y) \propto \exp\left(-\frac{1}{2}\langle y, (\alpha^2(2J)^{-1})y\rangle + \log(Z(\alpha y))\right).$$

Let $X$ be the random vector with density $\frac{f}{\int f}$. Then the above formula and the law of total variance give

$$\begin{aligned}\mathrm{Cov}(\nu) &= \mathrm{Cov}(\mathbb{E}[\mathcal{T}_{v+\alpha X}\nu]) + \mathbb{E}[\mathrm{Cov}(\mathcal{T}_{v+\alpha X}\mu)] \\ &= \mathrm{Cov}(\tanh(v+\alpha X)) + \mathbb{E}[\mathrm{Cov}(\mathcal{T}_{v+\alpha X}\mu)] \\ &\preceq \alpha^2\mathrm{Cov}(X) + \mathbf{I}_n, \end{aligned} \tag{79}$$

where we used the fact that $\tanh(X)$ is a contraction.

Since $\log Z(\cdot)$ is the logarithmic Laplace transform of the uniform measure, we obtain that $\nabla^2 \log(Z(w)) \preceq \mathbf{I}_n$. Therefore, defining

$$U(y) = \frac{1}{2}\langle y, (\alpha^2(2J)^{-1})y\rangle - \log(Z(\alpha y)),$$

we have

$$\nabla^2 U \succeq \alpha^2(2J)^{-1} - \alpha^2\mathbf{I}_n \succeq \left(\frac{\alpha^2}{2\|J\|_{\mathrm{OP}}} - \alpha^2\right)\mathbf{I}_n.$$

An application of Theorem 57 gives

$$1 + \|\alpha^2\mathrm{Cov}(X)\|_{\mathrm{OP}} \leq 1 + \frac{2\|J\|_{\mathrm{OP}}}{1-2\|J\|_{\mathrm{OP}}} = \frac{1}{1-2\|J\|_{\mathrm{OP}}}.$$

Combining with (79) completes the proof. □

# C Appendix C: The hardcore model

*Proof of Lemma 64.* Denote $\mathfrak{E}(\sigma)$ the event $\sigma_i u_i \geq 0, \forall i \in [n]$. For the upper bound, we have

$$\begin{aligned}\mathbb{P}_{\sigma\sim\nu}\left(\sigma_v = +1 \mid \mathfrak{E}(\sigma)\right) &= \frac{\sum_{\sigma_v=+1,\sigma_i u_i\geq 0,\forall i}\lambda^{|I_\sigma|}}{\sum_{\sigma_v=+1,\sigma_i u_i\geq 0,\forall i}\lambda^{|I_\sigma|} + \sum_{\sigma_v=-1,\sigma_i u_i\geq 0,\forall i}\lambda^{|I_\sigma|}} \\ &\overset{(i)}{\leq} \frac{\sum_{\sigma_v=+1,\sigma_i u_i\geq 0,\forall i}\lambda^{|I_\sigma|}}{\sum_{\sigma_v=+1,\sigma_i u_i\geq 0,\forall i}\lambda^{|I_\sigma|} + \sum_{\sigma_v=+1,\sigma_i u_i\geq 0,\forall i}\lambda^{|I_\sigma|-1}} \\ &= \frac{\lambda}{\lambda+1}.\end{aligned}$$

Inequality (i) follows from the observation that any configuration $\sigma$ with $\sigma_v = +1$ gives a configuration $\sigma'$ with $\sigma'_v = -1$ and $\sigma'_w = \sigma_w, \forall w \in V \setminus \{v\}$.
For the lower bound, let $(u_1, \ldots, u_K)$ be the elements of $N_v$, $K = |N_v|$, we have

$$\begin{aligned}&\mathbb{P}_{\sigma\sim\nu}\left(\sigma_v = +1 \mid \mathfrak{E}(\sigma)\right) \\ &\geq \mathbb{P}_{\sigma\sim\nu}\left(\sigma_v = +1 \text{ and } \sigma_{u_k} = -1, \forall k \in [K] \mid \mathfrak{E}(\sigma)\right) \\ &= \mathbb{P}_{\sigma\sim\nu}\left(\sigma_{u_k} = -1, \forall k \in [K] \mid \mathfrak{E}(\sigma)\right) \cdot \mathbb{P}_{\sigma\sim\nu}\left(\sigma_v = +1 \mid \sigma_{u_k} = -1, \forall k \in [K], \mathfrak{E}(\sigma)\right)\end{aligned}$$

Conditioned on all neighbors of $v$ being $-1$, $\sigma_v$ can either be $-1$ or $+1$. We have

$$\mathbb{P}_{\sigma\sim\nu}\left(\sigma_v = +1 \mid \sigma_{u_k} = -1, \forall k \in [K], \mathfrak{E}(\sigma)\right) = \frac{\lambda}{1+\lambda}.$$

On the other hand, we have

$$\begin{aligned}
&\mathbb{P}_{\sigma\sim\nu}\left(\sigma_{u_k} = -1, \forall k \in [K] \mid \mathfrak{E}(\sigma)\right) \\
&= \prod_{i=1}^{K} \mathbb{P}_{\sigma\sim\nu}\left(\sigma_{u_i} = -1 \mid \sigma_{u_j} = -1, \forall j \in [i-1], \mathfrak{E}(\sigma)\right) \\
&= \prod_{i=1}^{K} \left(1 - \mathbb{P}_{\sigma\sim\nu}\left(\sigma_{u_i} = +1 \mid \sigma_{u_j} = -1, \forall j \in [i-1], \mathfrak{E}(\sigma)\right)\right) \\
&\overset{(i)}{\geq} \prod_{i=1}^{K} \left(1 - \frac{\lambda}{1+\lambda}\right) \\
&= \left(\frac{1}{1+\lambda}\right)^{|N_v|}.
\end{aligned}$$

Inequality (i) follows from the upper bound of the marginal in the first part.
If in addition $\lambda \leq (1-\delta)\lambda_\Delta$, then we have

$$\begin{aligned}
\left(\frac{1}{1+\lambda}\right)^{|N_v|} &\geq \left(\frac{1}{1+\lambda_\Delta}\right)^{|N_v|} \\
&\geq \left(\frac{1}{1+\lambda_\Delta}\right)^{\Delta} \\
&\overset{(i)}{\geq} \left(\frac{1}{1+3e^2/\Delta}\right)^{\Delta} \\
&\overset{(ii)}{\geq} e^{-3e^2}.
\end{aligned}$$

Inequality (i) follows from the fact that for $\Delta \geq 3$,

$$\lambda_\Delta = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}} = \frac{1}{\Delta-2}\left(1 + \frac{1}{\Delta-2}\right)^{\Delta-1} \leq \frac{1}{\Delta-2} e^{\frac{\Delta-1}{\Delta-2}} \leq \frac{3e^2}{\Delta}.$$

Inequality (ii) follows from $\log(1+x) \leq x,$ for $x \in (-1, 1)$. □