Trickle-Down in Localization Schemes and Applications

Nima Anari[1], Frederic Koehler[2], and Thuy-Duong Vuong[1]

[1]Stanford University, {anari,tdvuong}@stanford.edu
[2]University of Chicago, fkoehler@uchicago.edu

Abstract

Trickle-down is a phenomenon in high-dimensional expanders with many important applications — for example, it is a key ingredient in various constructions of high-dimensional expanders or the proof of rapid mixing for the basis exchange walk on matroids and in the analysis of log-concave polynomials. We formulate a generalized trickle-down equation in the abstract context of linear-tilt localization schemes. Building on this generalization, we improve the best-known results for several Markov chain mixing or sampling problems — for example, we improve the threshold up to which Glauber dynamics is known to mix rapidly in the Sherrington-Kirkpatrick spin glass model. Other applications of our framework include improved mixing results for the Langevin dynamics in the $O(N)$ model, and near-linear time sampling algorithms for the antiferromagnetic and fixed-magnetization Ising models on expanders. For the latter application, we use a new dynamics inspired by polarization, a technique from the theory of stable polynomials.

Contents

1 Introduction

Recent years have seen an explosion of work on understanding *high-dimensional expansion*, in other words, higher-dimensional analogues of expansion in graphs. This has contributed to significant breakthroughs in our understanding of Markov chains, error-correcting codes, the geometry of polynomials, and other areas. See, e.g., [CGM19a; AL20; KO18; Ana+19b; DK17] for a few relevant references.

The present work is inspired by one of the key ingredients in the theory of high-dimensional expanders: Oppenheim's *trickle-down* (or trickling-down) theorem and the related Garland method [Opp18; Gar73]. Trickle-down expresses the following geometric idea: if a simplicial complex is locally very well-connected (in the sense of expansion), then the complex must be composed of disjoint pieces, each globally very well-connected.

To be a bit more precise, a pure simplicial complex of dimension $k-1$ with n vertices is described by a (possibly weighted) subset of $\binom{[n]}{k}$, which are the highest-dimensional faces of the complex. The "top links" of the complex are induced by conditioning these faces to contain a specific set of $k-2$ vertices and forming the family/distribution of the additional 2 vertices; these top links can be thought of as a graph whose edges are these 2-sized subsets. Oppenheim's trickle-down tells us that if the top links of the complex are very good spectral expanders, then this "trickles down" to ensure the complex itself is a spectral high-dimensional expander provided the mild condition that it is not disconnected at any level. Spectral expansion of the complex means that a natural $1 \to k \to 1$ *up-down walk*, which starts at a vertex, moves to an adjacent face, and then goes to a uniformly random vertex of that face, has a large spectral gap.

Important Markov chains like the Glauber dynamics can naturally be viewed as types of down-up walks on a weighted simplicial complex, and trickle-down has revealed itself to be an invaluable tool for proving sharp mixing time bounds for important classes of distributions like spanning trees, bases of matroids, more generally, distributions induced by log-concave polynomials, and quite recently edge colorings of graphs; see, e.g., [Ana+19b; Ana+20; CGM19a; ALG22].

Our contribution. In this work, we extend the reach of the trickle-down approach beyond the context of high-dimensional expanders. We do this by formulating a trickle-down equation in the more general context of *linear-tilt localization schemes*. Localization schemes, introduced by Chen and Eldan [CE22], constitute a framework that naturally generalizes the concept of iterative "pinning" in high-dimensional expanders (conditioning on vertices, as in the discussion of "links" above). In short, linear-tilt localization schemes generalize the operation of pinning to iterative reweighing of a measure by linear functionals (a.k.a. tilts). The generalized trickle-down equation gives a systematic way to perform backward induction and to show that the original measure is "well-connected" as long as the simpler/localized measures are. See Section 1.2 for more details.

The new perspective arising from the trickle-down equation allows us to make progress on several important problems in the sampling literature, and also derive important structural consequences like concentration of measure which have so far been beyond the reach of existing tools. We briefly overview a couple of representative applications here, but there are more — see Section 1.1 and Section 1.3.

Example application: mixing in the Sherrington-Kirkpatrick (SK) model. The SK model is a celebrated spin glass model from statistical physics [SK75; MPV87]. This is an Ising model where

the interaction matrix is sampled from the GOE (Gaussian Orthogonal Ensemble). In other words, the matrix J is a random $n \times n$ symmetric matrix with independent entries (for $i < j$)

$$J_{ij} \sim \mathcal{N}(0, \beta^2/n)$$

where $\beta \geq 0$ is the *inverse temperature*, and the induced probability measure ν on $\{\pm 1\}^n$ is given by

$$\nu(x) \propto \exp(\langle x, Jx \rangle / 2) .$$

It was previously known for $\beta < 0.25$ that, with high probability over the choice of J, the Glauber dynamics[1] mixes in nearly-linear time [BB19; EKZ21; Ana+21a]. This threshold does not have any apparent physical meaning (for example, there is no phase transition associated with the Gibbs measure until $\beta = 1$). However, mathematically it was an inherent limit of previous techniques because it is the sharp threshold for an associated functional, arising from the Hubbard-Stratonovich transform, to be *convex*. Analyzing the trickle-down equation for this model arising from stochastic localization, we can break the convexity barrier and still obtain optimal mixing times:

Theorem 1 (Informal)**.** *For the SK model, there exists absolute constant $c > 0$ such that up to $\beta = 0.25 + c \approx 0.295$, with high probability over the choice of J, Glauber dynamics mixes in $O(n \log n)$ time.*

This is obtained as a special case of a general result that improves the mixing time bounds for many Ising models in terms of the spectrum of their interaction matrix, Theorem 3.

Another feature of the generalized trickle-down is that it makes sense for continuous models as well as discrete ones. We obtain, analogously to the improvement in the aforementioned result for Ising models, an improvement for higher-spin-dimension variants, namely, the well-known $O(N)$ model, under both the Langevin and Glauber dynamics. See Section 1.3 for details.

Example application: mixing for antiferromagnetic Ising models on expander graphs. The *antiferromagnetic* Ising model at inverse temperature $\beta \geq 0$ with external field $h \in \mathbb{R}^n$ on a graph G is the probability measure on the hypercube $\{\pm 1\}^n$ given by

$$\nu(x) \propto \exp\left(-\beta \sum_{i \sim j} x_i x_j + \langle h, x \rangle \right)$$

where $i \sim j$ is the adjacency relation in graph G. For a *worst-case* graph, the largest value of β up to which polynomial time sampling is possible is known and it corresponds to what is called the "tree uniqueness threshold" for the Gibbs measure on the infinite d-regular tree [Wei06; MWW09; CLV21a]. The uniqueness threshold is asymptotically $\simeq 1/d$ as $d \to \infty$. We show this can be greatly improved in expander graphs.

Theorem 2 (Informal)**.** *Consider a d-regular graph with adjacency matrix A on n vertices, and suppose that $\max\{|\lambda_2(A)|, |\lambda_n(A)|\} \leq \lambda$. For all $\beta < 1/2\lambda$ there exists an algorithm, which we call the polarized walk, which samples from the antiferromagnetic Ising model in $\widetilde{O}(nd)$ time.*

For example, in a random d-regular graph this means we can handle β up to a threshold of $\Theta(1/\sqrt{d})$ instead of $\Theta(1/d)$. We also obtain results for Glauber dynamics when d is small. See Section 1.1 and Section 1.3 for more discussion.

[1] This is a Markov chain which at every step, picks a random site i and resamples the spin X_i according to the conditional law $\nu(\cdot \mid X_{\sim i})$. It is also known as the Gibbs sampler.

1.1 Our results

We now proceed to formally state the main results of this work, obtained based on the generalized trickle-down equation. Since these results apply to large classes of models, we give more specialized and concrete examples illustrating their application in the later Section 1.3.

Our main results involve systems of n spins, with each of the n spins taking values on the unit sphere $S^{N-1} = \{x \in \mathbb{R}^N \mid \|x\| = 1\}$. We use σ^{N-1} to denote the natural "uniform" probability measure on S^{N-1}. As a special case, the 0-dimensional unit sphere is the discrete set $\{\pm 1\}$, and we use $\mathrm{Ber}_\pm$ to denote the uniform distribution in that case. Our main results bound the covariance matrix $\mathrm{cov}(\nu)$ of the underlying distribution ν, from which approximate tensorization of entropy, abbreviated as ATE, (see Section 2.3) and appropriate mixing time results follow.

For a distribution μ on (a subset of) $\mathbb{R}^N$, such as σ^{N-1} or $\mathrm{Ber}_\pm$, and vector $w \in \mathbb{R}^N$, we use $\mathcal{T}_w \mu$ to denote the exponentially tilted distribution given by the following density (see Definition 33):

$$\frac{d\mathcal{T}_w \mu}{d\mu}(x) \propto \exp(\langle w, x \rangle).$$

Improved spectral condition for rapid mixing of Glauber dynamics in Ising models. A distribution ν on $\{\pm 1\}^n$ is an *Ising model* with interaction matrix J and external field h if

$$\nu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx \rangle + \langle h, x \rangle\right)$$

for all $x \in \{\pm 1\}^n$. Note that the diagonals of J are arbitrary; they do not affect the measure ν.

Theorem 3 (Theorem 54 below)**.** *Suppose that ν is an Ising model parameterized by some external field $h \in \mathbb{R}^n$ and interaction matrix J satisfying $J \succeq 0$. Suppose, without loss of generality, that the diagonal of J is constant, so there exists α such that $J_{ii} = \alpha \geq 0$. Let $\eta = \alpha / \|J\|_{\mathrm{op}} \in [0, 1]$. Then*

$$\|\mathrm{cov}(\nu)\|_{\mathrm{op}} \leq q_\eta(\|J\|_{\mathrm{op}})$$

where $q_\eta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ solves the Volterra integral equation

$$q_\eta(z) = r(\eta z) + \int_0^z q_\eta(y)^2 dy$$

with

$$r(t) = \mathbb{E}_{x \sim \mathrm{Ber}_\pm, g \sim \mathcal{N}(0,1)}\left[\mathrm{cov}\left(\mathcal{T}_{tx + \sqrt{t}g} \mathrm{Ber}_\pm\right)\right] = \mathbb{E}_{x \sim \mathrm{Ber}_\pm, g \sim \mathcal{N}(0,1)}\left[1 - \tanh^2\left(tx + \sqrt{t}g\right)\right] \leq 1.$$

Furthermore, ν satisfies approximate tensorization of entropy with constant at most

$$\exp\left(\int_0^{\|J\|_{\mathrm{op}}} q_\eta(z) dz\right).$$

Remark 4 (Spectral interpretation). As mentioned above, there is no a priori significance to the diagonal entries of J. For this reason, some works, including this one, use the convention that J is positive semidefinite without loss of generality. On the other hand, it is probably most common to parameterize the Ising model in terms of an interaction matrix with zero diagonal. Such a matrix will typically have both positive and negative eigenvalues because the mean of its spectral

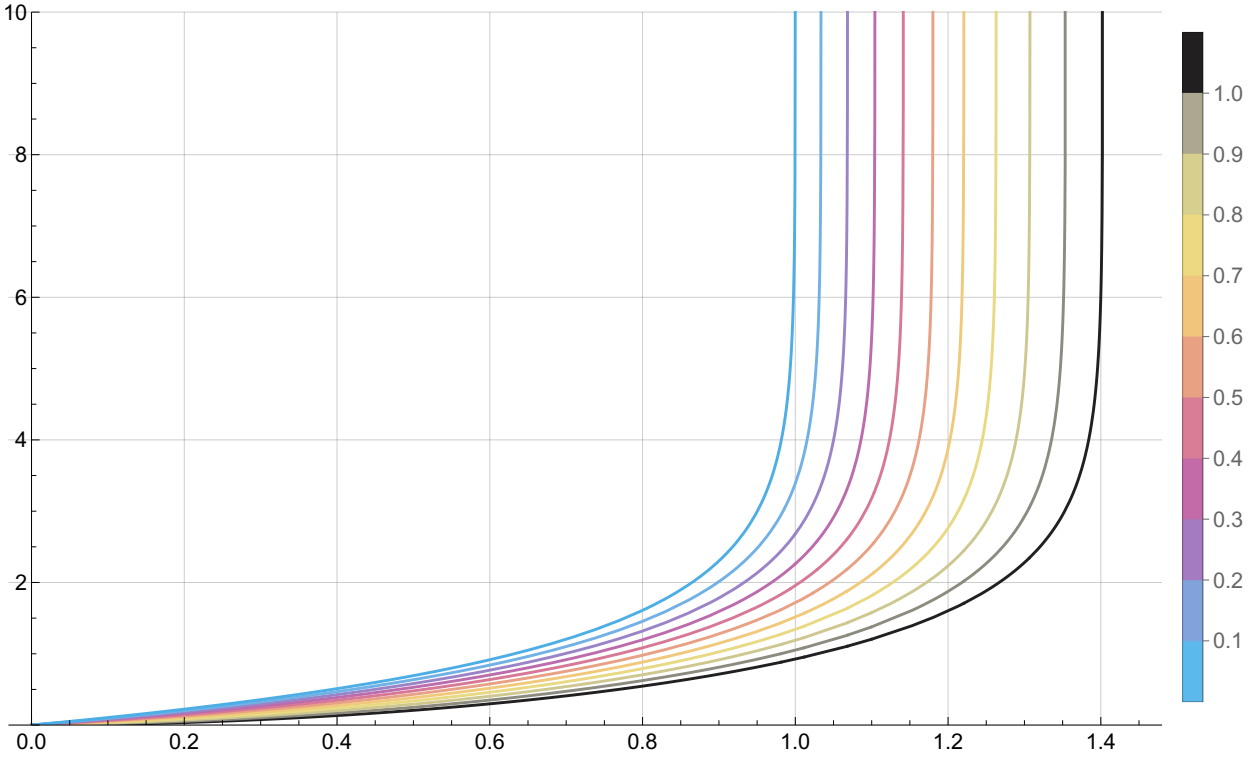

Figure 1: $\log q_\eta(z)$ from Theorem 3 plotted using numerical integration in Mathematica for $\eta = 0, 0.1, \ldots, 0.9, 1$. The curve for $\eta = 0$ was, before this work, the best known bound for all values of η in the Ising model; it asymptotes to ∞ at $z = 1$, which is tight due to the phase transition in the Curie-Weiss model [Ell06]. For $\eta = 0.5$ the function $\log q_\eta$ asymptotes to ∞ at z slightly above 1.18 and for $\eta = 1.0$ it asymptotes around 1.40.

distribution is zero. Fortunately, the above theorem has a very clean interpretation in terms of the zero-diagonal parameterization.

Let $\tilde{J}$ be a matrix with zero diagonals which specifies the interactions in the Ising model, so $\nu(x) \propto \exp(\langle x, \tilde{J}x\rangle/2 + \langle h, x\rangle)$. If we add a copy of the identity to $\tilde{J}$, we can always ensure the resulting matrix is positive semidefinite without changing the corresponding measure. More specifically, we can define $J = \tilde{J} + \alpha I$ where $\alpha = -\lambda_{\min}(\tilde{J})$. Then when we apply Theorem 3, our α will be the same as the one appearing in the theorem statement. We have

$$\|J\|_{\mathrm{op}} = \alpha + \lambda_{\max}(\tilde{J})$$

so

$$\eta = \frac{\alpha}{\|J\|_{\mathrm{op}}} = \frac{1}{1 + |\lambda_{\max}(\tilde{J})/\lambda_{\min}(\tilde{J})|} \in [0, 1].$$

We have $\eta \approx 0$ when the spectrum of $\tilde{J}$ has only very tiny negative eigenvalues. In that case, Theorem 3 reduces to the same bound as the existing results [BB19; EKZ21; Ana+21a]. However, in most applications, $\tilde{J}$ has non-negligible negative eigenvalues, so $\eta > 0$ and then this result significantly improves on the previous work — see Fig. 1 and examples in Section 1.3. For example, in many cases, $\tilde{J}$ has an approximately symmetrical spectrum and so $\eta \approx 1/2$.

Langevin dynamics in the $O(N)$ model. The $O(N)$ model [Sta68] is a famous class of models in statistical physics parameterized by an ambient dimension N. In this model, there are n interacting spins and each spin lives on the $(N-1)$-dimensional unit sphere S^{N-1}; the name $O(N)$ is a reference to symmetry under the N-dimensional orthogonal group. In the case $N = 1$, the 0-dimensional unit sphere is simply the points $\{\pm 1\}$ and so the $O(1)$ model is just the Ising model. The case $N = 2$ is the XY model and the case $N = 3$ is the (classical) Heisenberg model, which are very famous models in statistical physics; for example, the XY model on a two-dimensional lattice exhibits the celebrated BKT phase transition [KT18; Ber71] for which the 2016 Nobel Prize in Physics was awarded.

When $N \geq 2$ the unit sphere becomes connected, so the (manifold) Langevin dynamics is a natural sampling algorithm to analyze. Bauerschmidt and Bodineau [BB19] proved a log-Sobolev inequality for this dynamics under a bounded operator norm assumption on the interaction matrix J. Analogous to the Ising case, we prove a new result with a refined spectral condition that yields improved bounds in most applications. An interesting feature of this result is that for each N, the bound is given by an explicit rational function.

Theorem 5 (Theorem 72 below)**.** *Suppose $N \geq 2, n \geq 1$ and that ν is the probability distribution on $(S^{N-1})^n$ with probability density*

$$\frac{d\nu}{d\mu}(x) \propto \exp\left(\frac{1}{2}\sum_{1\leq i,j\leq n} J_{ij}\langle x_i, x_j\rangle + \sum_{i=1}^{n}\langle h_i, x_i\rangle\right)$$

with respect to the uniform measure $\mu = (\sigma^{N-1})^n$ on $(S^{N-1})^n$. Here J and h are parameters and we assume the interaction matrix J satisfies $J \geq 0$. Suppose, without loss of generality, that the diagonal of J is constant, so there exists α such that $J_{ii} = \alpha \geq 0$. Let $\eta = \alpha/\|J\|_{\mathrm{op}} \in [0, 1]$. Then[2]

$$\|\mathrm{cov}(\nu)\|_{\mathrm{op}} \leq q_{\eta,1/N}(\|J\|_{\mathrm{op}})$$

[2]This result is improved compared to the conference version of this paper [AKV24].

where $q_{\eta,1/N} = (1/N)Q_\eta(z/N)$ *and* $Q_\eta : [0, s(\eta)) \to \mathbb{R}_{\geq 0}$ *is given by*

$$Q_\eta(z) = \eta \frac{-\lambda_1\lambda_2^2(\eta z+1)^{\lambda_1-\lambda_2} + \lambda_1^2\lambda_2}{\lambda_2^2(\eta z+1)^{\lambda_1-\lambda_2+1} - \lambda_1^2(\eta z+1)}, \tag{1}$$

with $\lambda_1 > 0 > \lambda_2$ *the two roots of the quadratic equation* $\eta\lambda^2 - \eta\lambda - 1 = 0$*, explicitly*

$$\lambda_1 = \lambda_1(\eta) = \frac{\eta + \sqrt{\eta^2+4\eta}}{2\eta}, \qquad \lambda_2 = \lambda_2(\eta) = \frac{\eta - \sqrt{\eta^2+4\eta}}{2\eta},$$

and where

$$s(\eta) = \frac{1}{\eta}\left[\left|\frac{\lambda_1}{\lambda_2}\right|^{2/(\lambda_1-\lambda_2)} - 1\right].$$

Furthermore, the log-Sobolev constant of the Langevin dynamics is at least

$$C_N \exp\left(-\int_0^{\|J\|_{\mathrm{op}}} q_{\eta,1/N}(z)dz\right)$$

where $C_N > 0$ *is a constant depending only on* N*, inherited from [ZQM11].*

Example 6. The equations simplify considerably in the case $\eta = 1/2$: we have that $\lambda_1 = 2$ and $\lambda_2 = -1$,

$$Q_{1/2}(z) = \frac{-(z/2+1)^3 - 1}{(z/2+1)^4 - 2(z/2+1)} = \frac{48 + 24z + 12z^2 + 2z^3}{48 - 24z^2 - 8z^3 - z^4},$$

and

$$s(1/2) = 2\left[4^{1/3} - 1\right] \approx 1.1748.$$

Remark 7. Our methods automatically yield an additional result, Theorem 68, which for each particular value of N can likely yield a small improvement in the threshold obtained by Theorem 64. We expect the improvement is relatively small, and determining the quantitative guarantee requires a nontrivial computation for each value of N, so we did not further pursue this direction.

As stated, this result applies to the continuous-time (manifold) Langevin dynamics. To get an algorithm, some discretization scheme is usually applied. Fortunately, the log-Sobolev inequality also applies polynomial time guarantees for the discretized Langevin dynamics on manifolds [LE20].

Glauber dynamics in the $O(N)$ model: nearly linear time sampling. Besides the Langevin dynamics, there is another very natural dynamics to consider in the $O(N)$ model — the Glauber dynamics.

Compared to Langevin, Glauber dynamics has the advantage that it doesn't require discretization or choice of step size to implement. All known discretization schemes for the Langevin dynamics incur extra dimension-dependent factors, even in the simplest Euclidean setting [see, e.g., LST21]. Concretely, this means that while we established a "dimension-free" LSI constant, combining it with the discretization analysis of Langevin as in [LE20] pays an extra dimension factor of n, yielding a runtime guarantee of $\widetilde{O}(n^2 d)$ for a graphical model of maximum degree d (i.e., row-sparsity d in J) even in the simplest case $N = 2$.

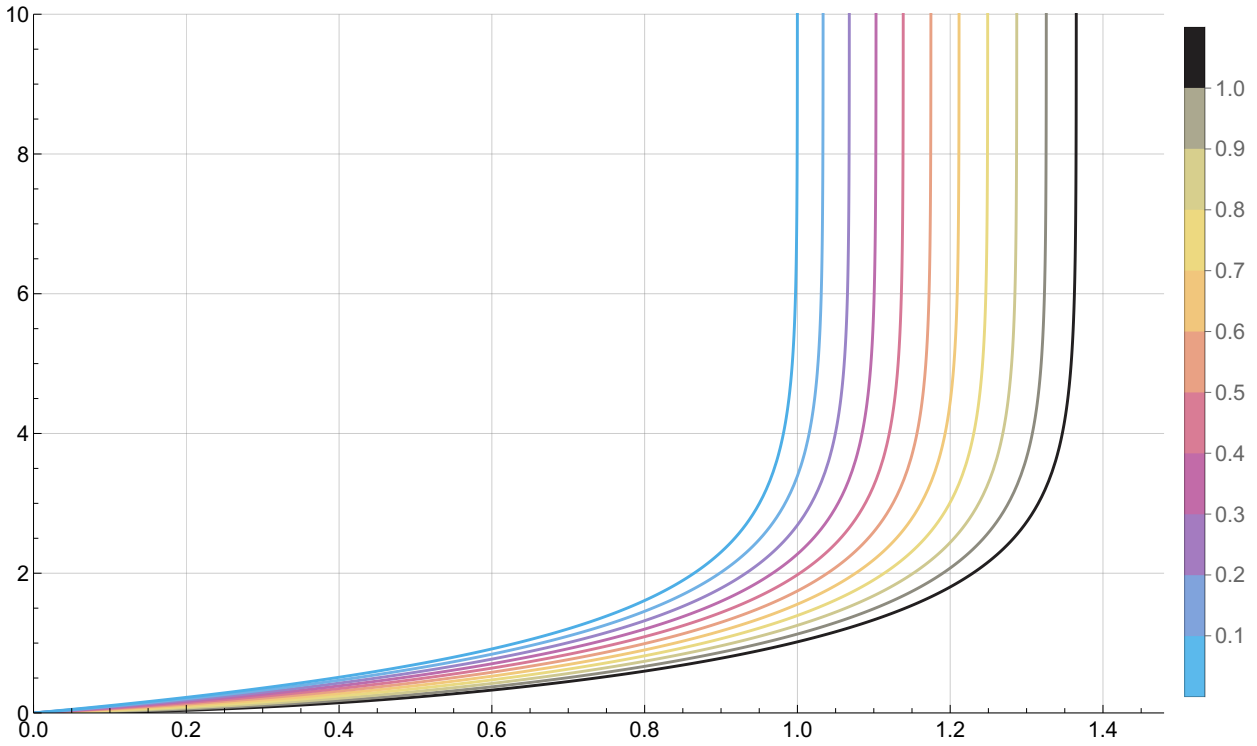

Figure 2: $\log Q_\eta(z)$ from Theorem 5 plotted using numerical integration in Mathematica for $\eta = 0, 0.1, \ldots, 0.9, 1$. For the $O(N)$ model, the actual function bounding the covariance is $q_{\eta,1/N}(z) = (1/N)Q_\eta(z/N)$. So for example, in the case of the XY model which is $N = 2$, the $\eta = 0.5$ curve asymptotes to ∞ slightly above 2.34 and for the Heisenberg model it asymptotes to ∞ slightly above 3.52. Similarly for $\eta = 1$, when $N = 2$ the resulting curve asymptotes to ∞ slightly above 2.73 for XY and at around 4.09 for Heisenberg.

We improve on this by proving that under the same conditions as before, the mixing time of Glauber dynamics is $O(n \log n)$, i.e., as fast as possible, and that each update step of the dynamics can be implemented very quickly. Together, these yield a nearly linear time sampler for our class of models. Here is the precise result:

Theorem 8 (Theorem 73 below)**.** *In the same setting as Theorem 5, the probability measure ν satisfies approximate tensorization of entropy (ATE) with constant*

$$\exp\left(\int_0^{\|J\|_{\mathrm{op}}} q_{\eta,N}(z)dz\right).$$

Furthermore, single steps of Glauber dynamics can be ϵ-approximately implemented using $O(dN + \log(1/\epsilon) \cdot (\log N + \log\log(1+B))$ arithmetic operations assuming at most d nonzeros in the corresponding row of J, and where $B = \max_i \sum_j |J_{ij}| + \|h_i\|$. In other words, we can sample from a distribution that is ϵ-close in total variation distance to the conditional distribution at the spin chosen to be updated. To sample from a distribution ε-close to ν in total variation distance, we run the Glauber dynamics for $T = n\log(n/\varepsilon)$ steps and set $\epsilon = \frac{\varepsilon}{T}$, resulting in an overall runtime of $\max\{\mathrm{nnz}(J), n\} \cdot N \cdot \mathrm{poly}\log(1/\varepsilon, n, N, \log(1+B))$.

Glauber dynamics over semi-log-concave base measures. In the mathematical physics literature, it is common to study more general distributions such as "soft spin" systems where the spin is $\mathbb{R}$-valued. See e.g. [SG73] for some references and discussion in the context of field theory. Our methods can give a general bound in terms of semi-log-concavity properties of the base measure of the spin system. We say that a measure is β-semi-log-concave if $\mathrm{cov}(\mathcal{T}_w \mu) \preceq \beta I$ for every $w \in \mathbb{R}^n$ [ES22]; see the preliminaries for more discussion.

Theorem 9 (Theorem 64 below)**.** *Suppose that $\mu = \bigotimes_{i=1}^n \mu^{(i)}$ is a product measure where each $\mu^{(i)}$ is a probability measure on $\mathbb{R}^N$. Let $J \succeq 0$ with $J_{ii} = \alpha$ for all $i \in [n]$, and define the measure ν by its density*

$$\frac{d\nu}{d\mu}(x) \propto \exp\left(\frac{1}{2}\sum_{i,j} J_{ij}\langle x_i, x_j\rangle\right).$$

Let $\eta = \alpha/\|J\|_{OP}$. Suppose that for all $i \in [n]$ and $s \in [0,\alpha]$ the measure $\mathcal{G}_{-s}\mu^{(i)}$ defined by

$$\frac{d\mathcal{G}_{-s}\mu^{(i)}}{d\mu^{(i)}}(x) \propto \exp(s\|x\|^2/2) \tag{2}$$

exists and is $\rho(s) > 0$ semi-log-concave. Then $\|\mathrm{cov}(\nu)\|_{OP} \le q_{\eta,\rho}(\|J\|_{OP})$ where $q_{\eta,\rho}$ solves the integral equation

$$q_{\eta,\rho}(z) = \frac{1}{\eta z + [\rho(\eta z)]^{-1}} + \int_0^z q_{\eta,\rho}(s)^2 ds \tag{3}$$

and ATE holds with constant at most

$$\exp\left(\int_0^{\|J\|_{OP}} q_{\eta,\rho}(z)dz\right).$$

The covariance bounds we obtain for the $O(N)$ model are a special case of this result — when $\rho(s) = \rho$ is constant, the integral equation admits an explicit solution $q_{\eta,\rho}(z) = \rho Q_\eta(\rho z)$ where Q_η is from Eq. (1). See Theorem 69.

Nearly linear time sampling for a class of Ising models via polarized walks. Antiferromagnetic Ising models are those where neighboring spins repel rather than attract each other. An impressive line of work has shown that for the antiferromagnetic Ising model on a worst-case d-regular graph, there is a computational phase transition at the uniqueness threshold of the infinite d-regular tree; see, e.g., [SS12; Wei06; GŠV16] for a few of the relevant works. This means that above a certain precise threshold for the strength of interactions in the model (inverse temperature), no polynomial time sampling algorithm exists under standard hypotheses from computational complexity.

What about non-worst-case graphs? Based on the picture coming from statistical physics as well as related rigorous results, we do not expect the uniqueness threshold to be relevant for random graphs; see [Coj+22] and references within. In particular, for d-regular random graphs sampling should be possible up to a threshold of $\Theta(1/\sqrt{d})$ instead of the $\Theta(1/d)$ scaling of the uniqueness threshold. For constant degrees d, this was recently proven by Koehler, Lee, and Risteski [KLR22], where a polynomial time algorithm inspired by variational inference was developed. This was recovered as part of a general result improving sampling guarantees for all expander graphs.

Can we improve this result to hold for all degrees d, to run in nearly linear time, and to be achieved by a simple Markov chain? Our Theorem 3, while applicable to antiferromagnetic Ising models, is not well-adapted to this setting because adjacency matrices of expander graphs have a large "trivial" eigenvalue, and its effect on the model is not so trivial.

Nevertheless, there is a natural class of models we can analyze using our techniques which captures both the antiferromagnetic Ising model on expanders as well as some other important applications like fixed-magnetization models. This class of models can be understood as spectrally bounded Ising models with an arbitrarily strong *confining potential* proportional to $(\sum_i x_i - k)^2$ for any $k \in \mathbb{Z}$. This potential[3] allows the measure to localize near a slice $\{x \mid \sum_i x_i \approx k\}$. Since this class of models includes those with fixed magnetization, Glauber dynamics may not even be ergodic. Fortunately, an interesting connection to the geometry of polynomials suggests a different Markov chain which does rapidly mix! We call this chain the *polarized walk* and describe it in more detail in Section 1.2.

Concretely, we establish the following result, which tells us that the polarized walk samples successfully in *nearly linear time* from a large class of models:

Theorem 10 (Theorem 86 below)**.** *Suppose that ν is a measure on the hypercube $\{\pm 1\}^n$ of the form*

$$\nu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle - \frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right)$$

for some $J \succeq 0$, $\gamma \geq 0$, and $h \in \mathbb{R}^n$, and suppose that $\|J\|_{\mathrm{op}} < 1/2$. Then:

1. *The polarized down-operator has $\left(1 - \frac{1-2\|J\|_{\mathrm{op}}}{n}\right)$-entropy contraction with respect to ν.*
2. *Hence, the polarized walk on ν mixes within ϵ total variation distance in $O\left(\frac{1}{1-2\|J\|_{\mathrm{op}}} n \log(n/\epsilon)\right)$ steps.*
3. *Let*

$$Q = J - \frac{\gamma}{n}\mathbb{1}\mathbb{1}^\intercal$$

[3] In the actual statement, we do not include the parameter k explicitly, since it can be absorbed into the external field of the model.

and let d be the maximum number of non-diagonal nonzero entries in a row of Q. Each step of the projected down-up walk can be implemented in $O(d \log n)$ amortized time, so the polarized walk outputs a sample within ϵ total variation distance of ν in total runtime $O\left(\frac{1}{1-2\|J\|_{\mathrm{op}}} nd \log(n) \log(n/\epsilon)\right)$

Corollary 11 (Corollary 91 below)**.** *Let ν be the antiferromagnetic Ising model with parameter β on a random d-regular graph G. Suppose $0 \le \beta \le (1-\delta)/(8\sqrt{d-1})$ for a constant $\delta > 0$. With probability $1-o(1)$ over the random instance G, we can sample from $\hat{\nu}$ s.t. $d_{\mathrm{TV}}(\hat{\nu}, \nu) \le \epsilon$ in time $O(\delta^{-1} nd \log(n) \log(n/\epsilon))$.*

Glauber dynamics on low-degree expanders. As discussed above, there is no hope of Glauber dynamics mixing for the general class of models studied in Theorem 10. However, it is still possible Glauber mixes rapidly for some subclasses of models. In the case of antiferromagnetic Ising models of bounded degree on expander graphs, we prove that the Glauber dynamics does indeed mix in $O(n \log n)$ time. One proof of this result is based on ideas from [KLR22] and in fact, yields mixing for a larger class of models(see Theorem 103); in Remark 105 we explain a variant of the result which is obtained using a trickle-down approach instead.

Theorem 12 (Corollary 104 below)**.** *Suppose that*

$$\nu(x) \propto \exp\left(-\frac{\beta}{2}\langle x, Ax\rangle + \langle h, x\rangle\right)$$

where A is the adjacency matrix of a d-regular graph on n vertices and suppose that $\max\{|\lambda_2(A)|, |\lambda_n(A)|\} \le \lambda$. If $\lambda\beta \le 1 - 1/c$ for some $c > 0$, then ν satisfies approximate tensorization of entropy with constant at most $e^{ce^{O(\beta d)}}$ and the Glauber dynamics on ν mixes in $O\left(e^{ce^{O(\beta d)}} \cdot n \log n\right)$ steps.

1.2 Our techniques

In this section, we give a brief overview of some of the techniques in this paper. In this part, we largely focus on the motivating applications involving the Glauber dynamics and polarized walk in the Ising model.

Challenge: nonconvexity of the Hubbard-Stratonovich transform. In the continuous domain, there has been a long and mathematically deep study of sampling *log-concave* distributions, i.e., probability densities $\propto e^{f(x)}$ where f is a concave function. In comparison, no distribution supported on the hypercube $\{\pm 1\}^n$ or products of unit spheres (as in the $O(N)$ model) can be log-concave because its support is not even a convex set.

Nevertheless, many previous works [BL02; DLS78; BB19; EKZ21; CE22; Ana+21a; KLR22] have been able to analyze Ising and $O(N)$ models using the celebrated *Hubbard-Stratonovich (HS) transform* [Hub59]. There are different ways to view this trick; at its heart, it essentially corresponds to applying the Fourier transform. However, a more probabilistic view is to say that we take $X \sim \nu$ for ν an Ising model with interaction matrix J and define a random vector

$$Y = X + G, \qquad G \sim \mathcal{N}(0, \Sigma), \qquad \Sigma = J^{-1}.$$

For this carefully chosen covariance matrix Σ, it can be checked via the Bayes rule that the posterior law $X \mid Y$ becomes a product measure. This is the Hubbard-Stratonovich trick. As a result, the task of sampling the *discrete random vector* X and *continuous random vector* Y become computationally equivalent. In particular, we can sample X and Y efficiently if the law of Y ends up being log-concave.

Moreover, more sophisticated analyses use the log-concavity of Y to derive rapid mixing of the Glauber dynamics for X.

This trick is elegant and it is *tight* in a certain sense: for the Curie-Weiss/mean-field Ising model,

$$p(x) \propto \exp\left(\frac{\beta}{2n}\left(\sum_i x_i\right)^2\right) \tag{4}$$

with $x \in \{\pm 1\}^n$ and inverse temperature $\beta \geq 0$, the high-temperature or rapid mixing regime of the model ($\beta \leq 1$) exactly corresponds to the set of parameters where the Hubbard-Stratonovich transform is log-concave.

However, it is less clear if this analysis is tight for other models. For the Sherrington-Kirkpatrick model, the Hubbard-Stratonovich transform is log-concave exactly when $\beta \leq 0.25$. Is this barrier fundamental? No. To go beyond this threshold (and prove all of our main results listed above), we need to develop new arguments which *do not rely on log-concavity.*

Remark 13. The papers [EKZ21; Ana+21a] use an alternative to the Hubbard-Stratonovich transform trick called *needle decomposition*, which decomposes the Ising model into Ising models with rank-one interaction matrices [cf. analogous notions in convex geometry, KLS95]. Although this is a different decomposition, it runs into the same barrier — the Curie-Weiss model is a rank-one model and the HS transform analysis is tight for it. The works [Adh+22; Ana+23] use an inductive approach instead, but only obtain results up to comparatively small values of β.

Trickle-down equation in localization schemes. The way we go beyond the log-concavity threshold is by finding a more "intrinsic" approach to analyze the mixing time of discrete distributions. From a long line of previous work [see, e.g., KO18; CGM19b; AL20; ALO20; Ana+21a; Ali+21; CLV21b; Ana+21b; BD22; CE22], we know that there are deep connections between proving mixing time bounds and establishing control of the covariance matrix of a distribution *under arbitrary tilts or pinnings*. As a reminder, exponential tilts are reweighings of a distribution $p(x)$ by a factor proportional to $\exp(\langle w, x\rangle)$ for some external field vector w. See also Section 2.4 of the preliminaries. So to prove our results, we develop a systematic method to control the covariance matrix of all of these tilts.

To do this, we take inspiration from the celebrated *trickle-down phenomenon* in high-dimensional expanders [Opp18], and develop a general trickle-down method that applies in the abstract setting of *linear-tilt localization schemes*. A linear-tilt localization scheme [CE22] for a probability measure ν_0 induced by a martingale difference sequence Z_t corresponds to a random sequence of measures

$$\frac{d\nu_{t+1}}{d\nu_t}(x) = 1 + \langle x - \mathrm{mean}(\nu_t), Z_{t+1}\rangle.$$

This general framework captures the usual pinning localization scheme used in high-dimensional expanders, stochastic localization, and other localization schemes like negative-field localization, etc. [CE22]. In this context, we prove the following generalized trickle-down equation (Eq. (8), see Section 3 for precise definitions and notation):

$$\mathrm{cov}(\nu_t) - \mathrm{cov}(\nu_t)\,\mathrm{cov}(Z_{t+1} \mid \mathcal{F}_t)\,\mathrm{cov}(\nu_t) = \mathbb{E}[\mathrm{cov}(\nu_{t+1}) \mid \mathcal{F}_t]$$

The key point is that this recursion allows us to perform backward induction over time to control the covariance matrix at time t by the covariance matrix at time $t+1$. In particular, we identify the exact analogue of trickle-down in stochastic localization, which we then use to prove our results.

Analyzing trickle-down via geometry of the base measure. A naïve application of the trickle-down method will only reprove the existing results obtained via the log-concavity of the Hubbard-Stratonovich transform. The reason for this is that for natural analogues of the distributions we study, these thresholds would actually be tight; we expand on this in the below example.

Example 14 (Gaussian SK model). For example, we could consider a "Gaussian SK model" which is defined just like the usual SK model but w.r.t. the Gaussian measure instead of the hypercube, so that its probability density function is

$$p(x) \propto \exp\left(\beta\langle x, Jx\rangle/2 - \|x\|_2^2/2\right),$$

where J is a scaled GOE matrix plus twice the identity, i.e., it is symmetric and independently $J_{ij} \sim \mathcal{N}(0, 1/n)$ for $i < j$ and $J_{ii} = 2$ for $i \in [n]$. The diagonal is included to ensure J is approximately PSD — this is a necessary convention because the stochastic localization process we use with trickle-down needs to use the driving matrix $J^{1/2}$. Both the base Gaussian measure $\propto e^{-\|x\|_2^2/2}dx$ in this example and the uniform measure on the hypercube have the same covariance, and in the Gaussian model the condition $\beta < 1/4$ is clearly sharp for rapid mixing. More precisely, the Gaussian model does not even exist when $\beta > 1/4$, since the top eigenvalue of J is concentrated about 4, and if $\beta J/2 - I$ has eigenvalues ≥ 0, then the integral $\int p(x)dx$ becomes infinite.

So to obtain any actual improvement for SK and other Ising models, we *need* to use something fundamental about the fact that our distribution is supported on the hypercube $\{\pm 1\}^n$. Ultimately, the improvement arises from specific concentration/anticoncentration properties which the Gaussian measure lacks. Here is an example, which helps illustrate the key role that the diagonal entries of J play in our analysis:

Lemma 15. *If $X \sim \nu$ where the probability measure ν on $\{\pm 1\}^n$ is the Ising model with interaction matrix $J \geq 0$ and external field h, then for any site $i \in [n]$ we have*

$$\mathbb{P}_\nu[|\langle J_i, X\rangle + h_i| \geq J_{ii}] \geq 1/2.$$

Proof. From the definition of the Ising model, we can observe that the conditional law at site i satisfies

$$\mathbb{E}_\nu[X_i \mid X_{\sim i}] = \tanh(\langle J_{i,\sim i}, X_{\sim i}\rangle + h_i),$$

where $X_{\sim i}$ denotes X with coordinate i removed. From this equation and the monotonicity of the tanh function, we see that X_i and $\langle J_{i,\sim i}, X_{\sim i}\rangle + h_i$ are positively correlated. So conditional on any value of $\langle J_{i,\sim i}, X_{\sim i}\rangle + h_i$, the random variable X_i has at least a 50% chance of having the same sign as this quantity, which proves the result. □

The reason this is useful is that the random vector $JX + h$ naturally arises as part of the external field for large times in the trickle-down procedure we use (see proof of Lemma 57; in fact, what we really do is observe a connection between a multi-spin and single-spin stochastic localization processes); the fact that it is large means the relevant product measure on the hypercube has a large bias, which means it has a small covariance. The actual anticoncentration analysis we use is more sophisticated and yields a tighter quantitative bound: see Section 4 for details, and Section 5 for more general results.

Polarized walks and the geometry of polynomials. Our general result for Glauber dynamics in the Ising model cannot handle models of the form studied in Theorem 10. This reflects a limitation of the Glauber dynamics itself — in the fixed-magnetization limit $\gamma \to \infty$ with $h = 0$, the distribution

becomes close to supported on the slice of the hypercube $\{x \mid \sum_i x_i = 0\}$ and the Glauber dynamics will not be ergodic. This is because every two points on the slice differ in at least two positions.

Fortunately, our analysis based upon the geometry of polynomials naturally suggests a different random walk, the *polarized walk*, which we show will sample in nearly linear time. We formally describe the dynamics in terms of its transition matrix in Section 7, but first explain the intuition behind it here. The ordinary Glauber dynamics can naturally be viewed as the composition of a down step (erasing the spin at a randomly chosen site) and an up step (resampling the erased spin). Since this is not ergodic for large values of γ, we instead use the following *polarized down step*:

1. Given as input state $x \in \{\pm 1\}^n$, select a coordinate i uniformly at random from $[n]$.
2. Output x with entry x_i set to -1.

Compared to the down step in the Glauber dynamics, which erases the spin at a particular site, the *polarized* down step is different in two key ways: first, it can only change coordinates of x which are equal to $+1$, and second, if it does pick such a coordinate i where $x_i = +1$, it sets the coordinate to -1 instead of erasing it.

Given this down step, the *polarized up step* takes as input a vector $y \in \{\pm 1\}^n$ and samples from the posterior distribution on $x \sim \nu$ assuming that y is the output of the projected down operator. This is completely analogous to the definition of the usual up step. Because the projected down step only replaces $+1$ spins by -1 spins, the projected up operator only replaces -1 spins by $+1$ spins. See Definition 84 for the explicit transition matrix of this operator. Note that the polarized walk does not treat $+1$ and -1 symmetrically, even though they are symmetrical in the definition of the model. This means that an alternative dynamics with $+1$ and -1 reversed can also be used.

Once we choose a way to break the symmetry, these dynamics naturally arise from a construction in the geometry of polynomials called *polarization*, which is a natural way to map a polynomial to a multiaffine polynomial while preserving the property of being *log-concave/Lorentzian* [Ana+21c; BH20]. These polynomials show up in our work by identifying a spin vector $x \in \{\pm 1\}^n$ with the set x_+ of indices with $+1$ spins, and then looking at the *generating polynomial* $\sum_i \mu(S) x^S$ of the corresponding probability measure μ on sets. See Section 2.5 for details.

By setting up the right stochastic localization process, we decompose our original Ising measure into a mixture of negatively-spiked rank one Ising models. As a remark, this is in contrast to [EKZ21; Ana+21a], where the decomposition used positively-spiked rank one models. More specifically, we obtain models of the form

$$p(x) \propto \exp\left(-\frac{\gamma}{2n}\left(\sum_i x_i\right)^2 + \langle h, x\rangle\right).$$

This can be thought of as the antiferromagnetic analogue of the classical Curie-Weiss model (Eq. (4)); interestingly, this simple-looking statistical physics model ends up connecting in a very elegant way with the geometry of polynomials. We use this connection to prove that the polarized dynamics mixes rapidly in this model, and then using trickle-down along with some new facts about entropy contraction in stochastic localization (see Lemma 39), we establish the mixing time bound in Theorem 10.

Quickly implementing the up step of the polarized dynamics is nontrivial — to do this, we build a data structure based on self-balancing binary search trees (see Proposition 87). This completes the construction of the nearly linear time sampling algorithm.

1.3 Some example applications

Sherrington-Kirkpatrick (SK) model Recalling the discussion earlier, in the SK model the matrix J is a random $n \times n$ symmetric matrix with independent entries (for $i < j$)

$$J_{ij} \sim \mathcal{N}(0, \beta^2/n)$$

where $\beta \geq 0$ is the *inverse temperature*. It was previously known for $\beta < 0.25$ that the Glauber dynamics mixes in nearly linear time [BB19; EKZ21; Ana+21a]. From the famous results of Wigner [see, e.g., AGZ10], we know that as $n \to \infty$, the smallest and largest eigenvalues of the zero-diagonal interaction matrix J are $-2 + o(1)$ and $2 + o(1)$ respectively. Hence, applying Theorem 3 to the equivalent interaction matrix $J - \lambda_{\min}(J) \cdot I$ with $\eta = 1/2 + o(1)$ (see Remark 4) shows that our general result implies nearly linear time mixing up to $\beta \approx 0.295$; see Fig. 1.

d-regular diluted SK model. In spin glass theory, *diluted* versions of mean-field spin glasses, which are supported on sparse random graphs, have long been studied [see, e.g., Tal10]; for example, with the motivation that sparse interactions are more realistic models of various physical phenomena. One natural diluted version of the SK model has its support on a random d-regular graph; for example, in the case of Rademacher disorder, for each edge (i, j) the edge weight J_{ij} would be sampled i.i.d. and uniformly from $\{\pm\beta\}$. In this case, the zero-diagonal interaction matrix has operator norm at most $2\sqrt{d-1} + o(1)$ with high probability by a version of Friedman's theorem [Des+19; Fri03]. Thus, the best previous results yield $O(n \log n)$ mixing time up to the threshold $\frac{0.25}{\sqrt{d-1}}$ and our result with $\eta = 1/2 + o(1)$ improves the guarantee to hold up to $\approx \frac{0.295}{\sqrt{d-1}}$.

Hopfield networks. Hopfield networks [Lit74; Hop82; PF77] are a neural model of associative memory that have been hugely influential and extensively studied. Formally, given i.i.d. random patterns $\eta_1, \ldots, \eta_m$ sampled uniformly from $\{\pm 1\}^n$, the Hopfield network at inverse temperature $\beta \geq 0$ is the Ising model with interaction matrix

$$J = \frac{\beta}{2n} \sum_{v=1}^{m} \eta_v \eta_v^\intercal .$$

This is thought of as a "Hebbian" learning rule because for each memory η_v and for "neurons" i and j, the term $(\eta_v)_i(\eta_v)_j$ is positive if $(\eta_v)_i = (\eta_v)_j$ and negative otherwise. Therefore if J is thought of as the "wiring" of the neurons, then for each pattern all of the neurons which "fire together", i.e., have the same spin, are "wired together".

When the number of memories m is a constant, there is a polynomial time sampling algorithm for this model for all β [KLR22]. But this is not the only regime of interest. Perhaps the most studied setting is when m and n jointly go to infinity at a fixed aspect ratio $m/n \simeq \lambda$. In random matrix theory, statistics, and other areas this is called a "proportional scaling limit", and in this limit the spectrum of the interaction matrix J will go to a scaled version of the celebrated *Marchenko and Pastur law* [MP67]. From the definition of the model, we know that the diagonal entries all equal $\frac{\beta m}{2n} = \frac{\beta\lambda}{2}$. This means the spectrum of the zero-diagonal matrix $\tilde{J} = J - \frac{\beta\lambda}{2} \cdot I$ is a shifted and scaled version of the Marchenko and Pastur law. For every value of λ, applying our main result will improve the threshold for rapid mixing compared to prior works [BB19; EKZ21; Ana+21a].

Unlike the previous examples, the optimal choice of η will vary depending on λ because the Marchenko and Pastur law is asymmetrical. The precise improvement will depend on the particular

value of λ. For example, when $\lambda = 1$, the largest eigenvalue of J will be $2\beta + o(1)$ asymptotically almost surely, and the diagonal entry will be $\beta/2$. Applying our result with $\eta \approx 0.25$ will improve the guaranteed rapid mixing threshold from around $\beta = 0.5$ to around $\beta \approx 0.54$.

$O(N)$ model with bounded row norms. The following discussion is related to a conjecture of Dyson, Lieb, and Simon [DLS78]; see Remark D.1 there. Suppose that the interaction matrix J has small row norms off the diagonal; more specifically, let

$$R = \|J\|_{\infty\to\infty} = \max\left\{ \sum_{j:j\neq i} |J_{ij}| \;\middle|\; i \in [n] \right\}.$$

We can assume the diagonal entries of J all equal R without loss of generality; then by Gershgorin's circle theorem, J is positive semidefinite with operator norm at most $2R$. Applying Theorem 5, we therefore obtain $O(n \log n)$ time mixing for all R up to $Ns_\eta(1/2)/2 \approx 0.5874N$. In general, for every value of N it was proved by Dyson, Lieb, and Simon [DLS78] that the covariance matrix has operator norm $O(1)$, i.e., satisfies a bound independent of the number of sites n, as long as $R < 0.5N$, and they conjectured that, at least in nice cases like ferromagnetic models on lattices, this bound can be improved to hold when $R < N$. Bauerschmidt and Bodineau [BB19] improved the covariance bound result to a log-Sobolev inequality up to the same threshold. Our result improves the threshold by an N-independent multiplicative constant without making any further assumptions on J.

Antiferromagnetic Ising on random graphs. For random d-regular graphs on n vertices, as long as $\beta < \frac{1-\delta}{8\sqrt{d-1}}$, we can sample from the antiferromagnetic Ising model with parameter β using the *polarized walk*; see Corollary 91. Our runtime is nearly linear in the number of edges, thus optimal up to log factors. In Corollary 104, we show that the Glauber dynamics mixes in $O\left(e^{O(\sqrt{d})} \cdot n \log n\right)$ steps. Our results also apply to other random graph models such as Erdős-Rényi.

Fixed magnetization (ferromagnetic or antiferromagnetic) Ising on random graphs. For random d-regular graphs on n vertices, as long as $\beta < \frac{1-\delta}{8\sqrt{d-1}}$, we can sample from the fixed magnetization (ferromagnetic or antiferromagnetic) Ising model with parameter β at an arbitrary magnetization level k using $O(\delta^{-1} k \log n)$ steps of the down-up walk (Corollary 96). Bauerschmidt, Bodineau, and Dagallier [BBD23] concurrently and independently gave a bound on the mixing time of the down-up walk in the same parameter regime, but their bound has an exponential dependency on the degree d, i.e., $\exp(O(\sqrt{d}))$; see Lemma 2.6 in [BBD23]. The main goal in [BBD23] is to bound the mixing time of the Kawasaki dynamics, a localized version of the down-up walk when one can only exchange the spin of neighboring vertices. They do so by comparing Kawasaki dynamics with the down-up walk, incurring an extra $\exp(O(\sqrt{d} + h)) \log^4 n$ factor [BBD23, Lemma 4.1] and thus showing that the Kawasaki dynamics mixes in $O(\exp(O(\sqrt{d}) + h) \cdot n \log^6 n)$ steps. The extra $\log^5 n \cdot \exp(O(\sqrt{d} + h))$ is unlikely to be necessary; our mixing time bound for the down-up walk is the first step to remove this dependency.

Using the equivalence between sampling and counting [JVV86], we have an FPRAS for the partition function of the fixed magnetization antiferromagnetic and ferromagnetic Ising at any magnetization level and up to the same threshold for β. The sum of these partition functions over all magnetization levels is precisely the partition function for the Ising model, thus we also obtain FPRAS for the partition function of Ising models, and thus algorithms to sample from these models using counting

to sampling reduction. This will work even in the case that the external field has inconsistent signs, which is #BIS-hard without further assumptions [GJ07]. In the ferromagnetic case and antiferromagnetic with fixed degree d cases, similar results were obtained in prior work of Koehler, Lee, and Risteski [KLR22] with a different algorithm; in the antiferromagnetic case, the polarized walk gives another possible algorithm with nearly linear runtime.

Our results also apply to other random graph models such as Erdős-Rényi (Corollary 97), and more generally to expanders (Proposition 95).

1.4 Other related work

There have been many works studying Glauber dynamics in continuous state spaces. For example, in convex bodies Glauber dynamics is the well-known "coordinate hit-and-run" Markov chain, see, e.g., [NS22; LV23]. Glauber dynamics has also been analyzed in spin systems [Wu06; Mar13] under conditions similar to Dobrushin's condition.

The recent works [AMS22; Cel22] (see also [MW23]) developed and analyzed a sampler for the SK model based on AMP (Approximate Message Passing) and stochastic localization. An advantage of this approach is that it provably works up to $\beta < 1$, which is the entire replica-symmetric regime of the model with zero external field. On the other hand, this method ends up sampling in a much weaker sense (achieving sublinear Wasserstein distance), does not apply to related models like diluted spin glasses, and does not imply structural properties of the measure like subgaussian concentration. It is worth noting that the work [AMS22] also shows an impossibility result for sampling via a class of Lipschitz algorithms in the replica symmetry-breaking regime ($\beta > 1$); it is possible that this is an impossible task for all polynomial time algorithms. Besides the SK model, there have also been recent works on mixing in spin glasses with higher order interactions (p-spin models) on the sphere and hypercube at sufficiently high temperature [GJ19; Adh+22; Ana+23].

Also specifically in the context of the SK model, recent works [AG22; Bre+22; BXY23] have obtained $O(1)$ bounds on the operator norm of the covariance matrix of the SK model when $\beta < 1$, improving polylogarithmic bounds due to Talagrand [Tal11]. To contrast with our analyses, we establish (and crucially need) bounds which *hold uniformly over all external fields* h for a fixed/"quenched" realization of the interaction matrix J. This is because an $O(1)$ bound on the operator norm of the covariance matrix is, by itself, obviously not a sufficient condition for rapid mixing, Lipschitz concentration, and so on to hold for a probability measure. Because of the difficulties involved in handling all external fields h at once, there is unfortunately no obvious way to apply the techniques based on the cavity method and TAP equations used in those works.

Recently, Kunisky [Kun23] gave some evidence that the constant in the spectral condition used to guarantee polynomial time mixing of the Glauber dynamics in [EKZ21; Ana+21a] is sharp not just for Glauber dynamics (where it is tight in the Curie-Weiss model) but in fact for all polynomial time algorithms. It would be interesting if anything along these lines can be said for the more sophisticated spectral conditions established in Theorem 3 and Theorem 64.

1.5 Organization

In Section 2, we state general facts and preliminaries which are used in the following sections. In Section 3, we formulate the general trickle-down equation for linear-tilt localization schemes, which is used in the remainder of the body of the paper. In Section 4, we prove Theorem 3 concerning rapid mixing of the Glauber dynamics in Ising models. In Section 5 we prove the analogous results

with semi-log-concave base measures. In Section 6, we apply the semi-log-concave theory to obtain our results for both the Langevin and Glauber dynamics in the $O(N)$ model. Section 7 proves rapid mixing of the polarized walk. Finally, in Appendix B we prove rapid mixing of the Glauber dynamics on low-degree expanders.

Acknowledgment

This work was supported by NSF CAREER Award CCF-2045354. Thuy-Duong Vuong was supported by a Microsoft Research Ph.D. Fellowship. Frederic Koehler was supported in part by NSF award CCF-1704417, NSF award IIS-1908774, and N. Anari's Sloan Research Fellowship, and thanks the participants of the Simons Institute program on Probability, Geometry, and Computation in High Dimensions for related discussions.

2 Preliminaries

In this section, we summarize, in a mostly self-contained fashion, the relevant facts about Markov chains, localization schemes, geometry of polynomials, etc., which are needed to prove our results. Some of the facts we state — in particular, Lemma 29 — appear to be more general than what has previously appeared in the literature, and we will need this level of generality.

We use $[n]$ to denote the set of n elements $\{1, \dots, n\}$. We also use $[\bar{n}]$ to denote $\{\bar{1}, \dots, \bar{n}\}$, where we think of $\bar{i}$ as an element distinct from i. We use $2^{[n]}$ to denote the family of subsets of $[n]$ and $\binom{[n]}{k}$ to denote subsets of size k. We use $\mathbb{1}$ to denote the all-ones vector, where the dimension is inferred from context. For a set $S \subseteq [n]$, we use $\mathbb{1}_S \in \mathbb{R}^n$ to denote the indicator vector of the set S. We use $\mathbb{1}_1, \mathbb{1}_2, \dots, \mathbb{1}_n$ to denote the standard basis vectors in $\mathbb{R}^n$.

By default, we treat vectors as column vectors. For vectors $u, v \in \mathbb{R}^n$ we use $\langle u, v \rangle \in \mathbb{R}$ to denote the inner product and $u \otimes v = uv^\intercal \in \mathbb{R}^{n \times n}$ to denote the outer product. We use $u^{\otimes 2}$ to abbreviate $u \otimes u = uu^\intercal$.

For a probability measure μ on a space Ω and function $f : \Omega \to \mathbb{R}^d$ we use the shorthands $\mathbb{E}_\mu[f]$ and μf to denote the expectation $\mathbb{E}_{x \sim \mu}[f(x)]$.

Definition 16 (Mean and covariance)**.** Suppose that μ is a probability measure on (a subset of) $\mathbb{R}^n$. We use $\mathrm{mean}(\mu)$ to denote $\mathbb{E}_{x \sim \mu}[x]$. Similarly, we define the covariance matrix as follows:

$$\mathrm{cov}(\mu) = \mathbb{E}_{x \sim \mu}[x^{\otimes 2}] - \mathbb{E}_{x \sim \mu}[x]^{\otimes 2}.$$

With some abuse of notation, for a random variable X, we use $\mathrm{mean}(X)$ and $\mathrm{cov}(X)$ to denote the mean and covariance of the law of X. We also naturally allow conditioning in mean and cov, which result in vector/matrix-valued random variables. So for a σ-algebra $\mathcal{F}$ and random variable X, we have $\mathrm{cov}(X \mid \mathcal{F}) = \mathbb{E}[X^{\otimes 2} \mid \mathcal{F}] - \mathbb{E}[X \mid \mathcal{F}]^{\otimes 2}$.

We use $\mathcal{N}(m, \Sigma)$ to denote the Gaussian distribution with mean m and covariance matrix Σ. We use $\mathrm{Ber}(p)$ to denote the Bernoulli distribution on $\{0, 1\}$ having mean p and use the shorthand Ber to denote $\mathrm{Ber}(1/2)$. We use $\mathrm{Ber}_\pm(m)$ to denote the distribution on $\{\pm 1\}$ whose mean is m and use the shorthand $\mathrm{Ber}_\pm$ to denote $\mathrm{Ber}_\pm(0)$.

2.1 Markov kernels

Definition 17 (Markov kernel)**.** A Markov kernel κ from Ω to Ω' assigns to every point $x \in \Omega$ a probability measure $\kappa(x, \cdot)$ on Ω'.

We assume Ω and Ω' are measurable spaces equipped with σ-algebras $\mathcal{F}, \mathcal{F}'$, which we usually omit for the sake of brevity. The kernel κ would be a function $\Omega \times \mathcal{F}' \to [0, 1]$ such that for each x, $\kappa(x, \cdot)$ is a probability measure on $(\Omega', \mathcal{F}')$, and for each $S \in \mathcal{F}'$ the function $x \mapsto \kappa(x, S)$ is measurable w.r.t. $(\Omega, \mathcal{F})$.

In the literature, sometimes Markov kernels are called channels. For finite spaces Ω, Ω', we view κ as a row-stochastic matrix. Note that when $\Omega' = \Omega$, κ could be viewed as defining the transitions of a Markov chain on Ω.

A Markov kernel κ from Ω to Ω' can be combined with a Markov kernel κ' from Ω' to Ω'' in order to give a Markov kernel $\kappa\kappa'$ from Ω to Ω'':

$$\kappa\kappa'(x, S) = \int_{y \in \Omega'} \kappa'(y, S)\kappa(x, dy) = \mathbb{E}_{y \sim \kappa(x, \cdot)}[\kappa'(y, S)].$$

For finite spaces, this corresponds to matrix multiplication. We can also take the direct product of two kernels κ and γ from Ω to Ω' and from Ω to Ω'' respectively, and define $\kappa \times \gamma$ to be a kernel from Ω to $\Omega' \times \Omega''$ with $\kappa \times \gamma(x, \cdot) = \kappa(x, \cdot) \times \gamma(x, \cdot)$.

A measure μ on Ω can be combined with a Markov kernel κ from Ω to Ω' to get another measure $\mu\kappa$ on Ω':

$$\mu\kappa(S) = \int_{x \in \Omega} \kappa(x, S)\mu(dx).$$

For finite spaces, this corresponds to a vector-matrix multiplication, viewing μ as a row vector. Note that if μ is a probability measure, i.e., if $\mu(\Omega) = 1$, then so is $\mu\kappa$.

While distributions multiply on the l.h.s. of a Markov kernel, functions multiply on the r.h.s. Given a Markov kernel κ from Ω to Ω' and a measurable function $f : \Omega' \to \mathbb{R}^d$, define $\kappa f : \Omega \to \mathbb{R}^d$ as

$$\kappa f(x) = \int_{y \in \Omega'} f(y)\kappa(x, dy) = \mathbb{E}_{y \sim \kappa(x, \cdot)}[f(y)].$$

Definition 18 (Semidirect product)**.** Consider a distribution μ on Ω and a Markov kernel κ from Ω to Ω'. We can generate random variables X, Y, by first sampling $X \sim \mu$ and then, conditioned on X, sampling $Y \sim \kappa(X, \cdot)$. Note that the marginal distribution of Y is $\mu\kappa$. The joint distribution of (X, Y) is the semidirect product $\mu \rtimes \kappa$. We also use $\kappa \ltimes \mu$ to denote the distribution of (Y, X).

Definition 19 (Time-reversal)**.** We say that a Markov kernel κ' is the time-reversal of the Markov kernel κ w.r.t. the distribution μ if $\kappa'(Y, \cdot)$ is the (regular) conditional probability distribution of X given Y, when $(X, Y) \sim \mu \rtimes \kappa$. A shorthand definition is that μ and κ define the same joint distribution as $\mu\kappa$ and κ' define:

$$\mu \rtimes \kappa = \kappa' \ltimes (\mu\kappa)$$

Although time-reversal, being a conditional probability, is technically not necessarily unique, with slight abuse of notation, we use the following convenient shorthand to denote $\kappa'(y, \cdot)$:

$$\mu \mid \kappa \rhd y$$

In Bayesian terminology, μ would be a prior distribution, and y an observation obtained by passing a sample from μ through the channel κ; $\mu \mid \kappa \rhd y$ is simply the posterior distribution, given this observation. For finite spaces, we can write this posterior as

$$(\mu \mid \kappa \rhd y)(x) = \frac{\mu(x)\kappa(x,y)}{\mu\kappa(y)}.$$

If γ is another kernel from Ω to Ω'', then observations of κ and γ commute with each other:

$$(\mu \mid \kappa \rhd y) \mid \gamma \rhd z = (\mu \mid \gamma \rhd z) \mid \kappa \rhd y.$$

This is the posterior of $X \sim \mu$ given that independent observations $Y \sim \kappa(X, \cdot)$ and $Z \sim \gamma(X, \cdot)$ were $Y = y, Z = z$.

While a kernel κ allows us to map measures from Ω to measures on Ω', the densities of these measures are mapped according to the time-reversal kernel κ'.

Proposition 20. *Suppose that μ is a probability measure on Ω and κ is Markov kernel from Ω to Ω' whose time-reversal is κ'. If ν is another measure on Ω absolutely continuous w.r.t. μ, then*

$$\frac{d(\nu\kappa)}{d(\mu\kappa)} = \kappa' \frac{d\nu}{d\mu}.$$

Proof. For any test function f on Ω', we have

$$\begin{aligned}\mathbb{E}_{y\sim\mu\kappa}\left[\left(\kappa'\frac{d\nu}{d\mu}\right)(y)\cdot f(y)\right] &= \mathbb{E}_{y\sim\mu\kappa}\left[\mathbb{E}_{x\sim\kappa'(y,\cdot)}\left[\frac{d\nu}{d\mu}(x)\right]\cdot f(y)\right] = \mathbb{E}_{x\sim\mu}\left[\frac{d\nu}{d\mu}(x)\cdot \mathbb{E}_{y\sim\kappa(x,\cdot)}\big[f(y)\big]\right] \\ &= \int_{x\in\Omega} \mathbb{E}_{y\sim\kappa(x,\cdot)}[f(y)]\ \nu(dx) = \int_{y\in\Omega'} f(y)\ \nu\kappa(dy).\end{aligned}$$

This proves that $\kappa'\frac{d\nu}{d\mu}$ is the density of $\nu\kappa$ w.r.t. $\mu\kappa$. □

Definition 21 (Mixture decomposition)**.** A mixture decomposition of a distribution μ on Ω is a collection of distributions μ_θ indexed by $\theta \in \Omega'$, together with a distribution π on Ω', such that for all events $S \subseteq \Omega$, $\mu_\theta(S)$ is measurable in θ and $\mu(S) = \mathbb{E}_{\theta\sim\pi}[\mu_\theta(S)]$. We abbreviate this as

$$\mu = \mathbb{E}_{\theta\sim\pi}[\mu_\theta].$$

Mixture decompositions are in direct correspondence with Markov kernels from Ω to Ω'. Given a mixture decomposition we define a Markov kernel κ: first let the kernel κ' be $\kappa'(\theta, \cdot) = \mu_\theta$, and note that $\pi\kappa' = \mu$. Now let κ be the time-reversal of κ' w.r.t. π. Going the opposite direction, for any Markov kernel κ from Ω to Ω', we can let $\pi = \mu\kappa$, and κ' be the time-reversal of κ w.r.t. μ, and define $\mu_\theta = \kappa'(\theta, \cdot)$. We automatically get $\pi\kappa' = \mu$, i.e., $\mathbb{E}_{\theta\sim\pi}[\mu_\theta] = \mu$.

2.2 Entropy contraction

Definition 22 (φ-entropy and φ-divergence)**.** Given a convex function $\varphi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ and measure μ on space Ω, φ-entropy is defined as an operator that takes functions $f : \Omega \to \mathbb{R}_{\geq 0}$ and outputs

$$\mathrm{Ent}_\mu^\varphi[f] = \mathbb{E}_\mu[\varphi \circ f] - \varphi\big(\mathbb{E}_\mu[f]\big)\,.$$

We also define the related notion of a φ-divergence.[4] For a measure ν that has density $d\nu/d\mu$ w.r.t. μ we define the φ-divergence as

$$\mathcal{D}_\varphi(\nu \parallel \mu) = \mathrm{Ent}_\mu^\varphi\left[\frac{d\nu}{d\mu}\right].$$

Note that φ-entropy and φ-divergence are always nonnegative, by Jensen's inequality. Two special cases of φ give the familiar variance and entropy functionals:

Definition 23 (Variance and χ^2-divergence)**.** Specializing to $\varphi(x) = x^2$, we define variance as

$$\mathrm{Var}_\mu[f] = \mathrm{Ent}_\mu^\varphi[f] = \mathbb{E}_\mu[f^2] - \mathbb{E}_\mu[f]^2,$$

and the χ^2-divergence as

$$\chi^2(\nu \parallel \mu) = \mathcal{D}_\varphi(\nu \parallel \mu) = \mathrm{Var}\left[\frac{d\nu}{d\mu}\right].$$

Definition 24 (Relative entropy and KL-divergence)**.** Specializing to $\varphi(x) = x \log x$, we define the relative entropy as

$$\mathrm{Ent}_\mu[f] = \mathrm{Ent}_\mu^\varphi[f] = \mathbb{E}_\mu[f \log f] - \mathbb{E}_\mu[f] \log(\mathbb{E}_\mu[f]),$$

and the Kullback–Leibler (KL) divergence as

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) = \mathrm{Ent}_\mu\left[\frac{d\nu}{d\mu}\right].$$

Definition 25 (Entropy contraction)**.** We say a Markov kernel κ from Ω to Ω' has $(1-\rho)$ contraction of φ-entropy w.r.t. a distribution μ on Ω if for all measures ν absolutely continuous w.r.t. μ we have

$$\mathcal{D}_\varphi(\nu\kappa \parallel \mu\kappa) \le (1-\rho)\, \mathcal{D}_\varphi(\nu \parallel \mu),$$

or equivalently, assuming κ' is the time-reversal of κ w.r.t. μ, if for all density functions f on Ω,

$$\mathrm{Ent}_{\mu\kappa}^\varphi[\kappa' f] \le (1-\rho)\, \mathrm{Ent}_\mu^\varphi[f].$$

We will use the following useful characterization of shrinkage of φ-entropies under κ:

Proposition 26. *Suppose that ν is a measure absolutely continuous w.r.t. a probability measure μ on Ω, and κ is a Markov kernel from Ω to Ω'. Then*

$$\mathcal{D}_\varphi(\nu \parallel \mu) - \mathcal{D}_\varphi(\nu\kappa \parallel \mu\kappa) = \mathrm{Ent}_\mu^\varphi\left[\frac{d\nu}{d\mu}\right] - \mathrm{Ent}_{\mu\kappa}^\varphi\left[\frac{d(\nu\kappa)}{d(\mu\kappa)}\right] = \mathbb{E}_{y\sim\mu\kappa}\left[\mathrm{Ent}_{\mu|\kappa\triangleright y}^\varphi\left[\frac{d\nu}{d\mu}\right]\right].$$

Proof. Let $f = \frac{d\nu}{d\mu}$. If κ' is the time-reversal of κ w.r.t. μ, we have $\frac{d(\nu\kappa)}{d(\mu\kappa)} = \kappa' f$, so

$$\mathcal{D}_\varphi(\nu\kappa \parallel \mu\kappa) = \mathrm{Ent}_{\mu\kappa}^\varphi[\kappa' f] = \mathbb{E}_{\mu\kappa}[\varphi \circ \kappa' f] - \varphi(\nu\kappa(\Omega')) = \mathbb{E}_{\mu\kappa}[\varphi \circ \kappa' f] - \varphi(\nu(\Omega)),$$

where we used the fact that $\nu\kappa(\Omega') = \nu(\Omega)$. Similarly, we can write

$$\mathcal{D}_\varphi(\nu \parallel \mu) = \mathrm{Ent}_\mu^\varphi[f] = \mathbb{E}_\mu[\varphi \circ f] - \varphi(\nu(\Omega)) = \mathbb{E}_{y\sim\mu\kappa}[\mathbb{E}_{x\sim\kappa'(y,\cdot)}[\varphi \circ f]] - \varphi(\nu(\Omega)).$$

Subtracting we get

$$\mathcal{D}_\varphi(\nu \parallel \mu) - \mathcal{D}_\varphi(\nu\kappa \parallel \mu\kappa) = \mathbb{E}_{y\sim\mu\kappa}\left[\mathbb{E}_{x\sim\kappa'(y,\cdot)}[\varphi \circ f] - \varphi(\kappa' f(y))\right] = \mathbb{E}_{y\sim\mu\kappa}\left[\mathrm{Ent}_{\kappa'(y,\cdot)}^\varphi[f]\right],$$

which finishes the proof. □

[4] f-divergence is more commonly used in the literature, but for consistency with φ-entropies, we use φ-divergence.

Note that Proposition 26 implies this difference is ≥ 0 for convex φ, since Ent^{φ} is always ≥ 0. In other words, we always have (weak) contraction of φ-entropy with $\rho \geq 0$. This is known as the data processing inequality. Using Proposition 26 we can write $(1-\rho)$ contraction of φ-entropy in a useful form in terms of density functions:

Proposition 27. *Suppose μ is a probability measure on Ω and κ is a Markov kernel from Ω to Ω'. Then we have $(1-\rho)$ contraction of φ-entropy iff for all densities f,*

$$\mathbb{E}_{y\sim\mu\kappa}\left[\mathrm{Ent}^{\varphi}_{\mu|\kappa\triangleright y}[f]\right] \geq \rho\cdot\mathrm{Ent}^{\varphi}_{\mu}[f].$$

Remark 28. Entropy contraction for a Markov kernel κ and *all* probability measures μ is referred to as a strong data processing inequality [see, e.g., PW17]. In this work, we are concerned primarily with entropy contraction w.r.t. a fixed probability measure μ.

We now prove an important concavity property of the φ-entropy decrement.

Lemma 29. *Let κ be a Markov kernel from Ω to Ω', and let f be a density function on Ω. The following functional on distributions μ is concave:*

$$\mu \mapsto \mathbb{E}_{y\sim\mu\kappa}\left[\mathrm{Ent}^{\varphi}_{\mu|\kappa\triangleright y}[f]\right].$$

In other words, if we have a mixture decomposition for μ, given by $\mu = \mathbb{E}_{\theta\sim\pi}[\mu_\theta]$, then

$$\mathbb{E}_{y\sim\mu\kappa}\left[\mathrm{Ent}^{\varphi}_{\mu|\kappa\triangleright y}[f]\right] \geq \mathbb{E}_{\theta\sim\pi}\left[\mathbb{E}_{y\sim\mu_\theta\kappa}\left[\mathrm{Ent}^{\varphi}_{\mu_\theta|\kappa\triangleright y}[f]\right]\right].$$

Proof. A mixture decomposition can be described by a Markov kernel γ such that $\pi = \mu\gamma$, and $\mu_\theta = \mu \mid \gamma \triangleright \theta$. This means we can write the r.h.s. of the desired inequality as

$$\mathbb{E}_{\theta\sim\mu\gamma}\left[\mathbb{E}_{y\sim(\mu|\gamma\triangleright\theta)\kappa}\left[\mathrm{Ent}^{\varphi}_{(\mu|\gamma\triangleright\theta)|\kappa\triangleright y}[f]\right]\right] = \mathbb{E}_{y\sim\mu\kappa}\left[\mathbb{E}_{\theta\sim(\mu|\kappa\triangleright y)\gamma}\left[\mathrm{Ent}^{\varphi}_{(\mu|\kappa\triangleright y)|\gamma\triangleright\theta}[f]\right]\right].$$

Here we switched the order of expectation by using the fact that both sides take expectation w.r.t. $(y,\theta)\sim\mu\cdot(\kappa\times\gamma)$. For each y, by applying Proposition 26 to the distribution $\mu' = \mu \mid \kappa \triangleright y$ and Markov kernel γ, and using nonnegativity of Ent^{φ}, we get

$$\mathrm{Ent}^{\varphi}_{\mu'}[f] \geq \mathbb{E}_{\theta\sim\mu'\gamma}\left[\mathrm{Ent}^{\varphi}_{\mu'|\gamma\triangleright\theta}[f]\right].$$

Taking expectation w.r.t. $y\sim\mu\kappa$, and expanding $\mu' = \mu \mid \kappa \triangleright y$ finishes the proof:

$$\mathbb{E}_{y\sim\mu\kappa}\left[\mathrm{Ent}^{\varphi}_{\mu|\kappa\triangleright y}[f]\right] \geq \mathbb{E}_{y\sim\mu\kappa}\left[\mathbb{E}_{\theta\sim(\mu|\kappa\triangleright y)\gamma}\left[\mathrm{Ent}^{\varphi}_{(\mu|\kappa\triangleright y)|\gamma\triangleright\theta}[f]\right]\right]. \qquad \square$$

2.3 Functional inequalities and mixing

Glauber Dynamics and ATE. Our results for the Glauber dynamics go via establishing a fundamental information-theoretic inequality called *Approximate Tensorization of Entropy* (ATE). Let $\Omega = \Omega_1\times\cdots\Omega_n$ be a product space. For $i\in[n]$, define the Markov kernel $D_{\sim i}$ as one that maps $x = (x_1,\ldots,x_n)\in\Omega$ deterministically to $x_{\sim i}$ by erasing its ith component, i.e.,

$$x_{\sim i} \coloneqq (x_1,\ldots,x_{i-1},\varnothing,x_{i+1},\ldots,x_n).$$

Definition 30 (Approximate Tensorization of Entropy, [Mar13; CMT14])**.** Let μ be a probability measure on $\mathbb{R}^n$. We say μ satisfies Approximate Tensorization of Entropy (ATE) with constant C if for every other measure ν absolutely continuous w.r.t. μ,

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \leq C \sum_{i=1}^{n} \mathbb{E}_{x\sim\nu}\left[\mathcal{D}_{\mathrm{KL}}\big((\nu \mid D_{\sim i} \triangleright x_{\sim i}) \,\big\|\, (\mu \mid D_{\sim i} \triangleright x_{\sim i})\big)\right],$$

or equivalently, for all nonnegative functions f,

$$\mathrm{Ent}_{\mu}[f] \leq C \sum_{i=1}^{n} \mathbb{E}_{x\sim\mu}\left[\mathrm{Ent}_{\mu|D_{\sim i}\triangleright x_{\sim i}}[f]\right].$$

This inequality has a third equivalent reformulation. It asserts that

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{n\to n-1} \parallel \mu D_{n\to n-1}) \leq \left(1 - \frac{1}{Cn}\right) \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$$

where $D_{n\to n-1}$ is the Markov kernel that picks a coordinate $i \in [n]$ uniformly at random and erases it ("down operator"), i.e., the average of $D_{\sim 1}, \ldots, D_{\sim n}$. This can be seen from the equivalence described in Definition 25.

ATE is a very strong functional inequality, which yields an array of useful consequences "for free" once established. We do not go into full details about all of these consequences because we do not explicitly need them in this work, but give a brief overview below. They are fully explored in the references (with additional important references to prior work). First, we recall its implications for concentration of measure and mixing:

Proposition 31 (Fact 3.5 of [CLV21b])**.** *If ATE for measure μ holds with constant C then:*

1. *The MLSI (Modified Log-Sobolev Inequality) for Glauber dynamics holds with constant $1/Cn$.*
2. *As a consequence of the MLSI, on a discrete space, defining $\mu_{\min} = \min\{\mu(x) \mid x \in \mathrm{supp}(\mu)\}$, the ϵ-mixing time is $O(n(\log\log(1/\mu_{\min}) + \log(1/\epsilon)))$.*
3. *As a consequence of the MLSI, a W_1 entropy-transport inequality holds with constant C and this is equivalent to subgaussian concentration of Lipschitz functions with constant C and with respect to the Hamming metric.*
4. *As a consequence of the MLSI, the Poincaré inequality for Glauber dynamics holds with constant $1/Cn$.*
5. *If μ is defined on a discrete space and additionally b-marginally bounded, i.e., $\mathbb{P}_{X\sim\mu}[X_i = x_i \mid X_{\sim i} = x_{\sim i}] \geq b$ for all x, then the log-Sobolev inequality for Glauber dynamics holds with constant $O_b(1/Cn)$. This is furthermore equivalent to hypercontractivity of the semigroup.*

Here are some other useful consequences:

1. As a consequence of the MLSI, reverse hypercontractivity for the Glauber semigroup holds with an appropriate constant [MOS13].
2. As a consequence of the Poincaré inequality, μ^{hom} is C-spectrally independent [Ana+23]. Spectral independence is defined below in Section 2.5.
3. If ATE holds uniformly over all external fields (which is the case for all of the results in this paper), then spectral independence holds under all external fields by the previous point,

which is equivalent to fractional log-concavity of the generating polynomial (see Section 2.5). In turn, this implies entropic independence/subadditivity of entropy/Brascamp-Lieb type inequalities [CC09; Ana+21a].

4. Block tensorization of entropy holds with an appropriate constant — see [Lee23] for details.
5. ATE yields non-asymptotic bounds on the statistical performance of pseudolikelihood estimation [Bes75; KHR22].
6. ATE yields strong guarantees for identity testing in an appropriate oracle model [Bla+22].
7. For bounded degree and marginally bounded graphical models, ATE yields KKL and Talagrand isoperimetric inequalities [Koe+23].
8. On the hypercube $\{\pm 1\}^n$, ATE implies the log-Sobolev inequality for a different continuous-time dynamics: the inequality concretely takes on the form $\mathrm{Ent}[f] \lesssim C \sum_i \|\partial_i \sqrt{f}\|_2^2$ with $\partial_i g(x) = g(x_1, \ldots, x_{i-1}, +1, x_{i+1}, \ldots, x_n) - g(x_1, \ldots, x_{i-1}, -1, x_{i+1}, \ldots, x_n)$. The corresponding semigroup can be understood as a variant of Glauber with a variable transition rate (that can become much larger than one in some situations). This type of inequality was studied in, e.g.,[Mar99; BB19]; see [EKZ21] for more discussion.

Riemannian Langevin dynamics. Just like in Euclidean space, the Langevin diffusion on a Riemannian manifold is a continuous random walk. Given a probability distribution with relative density $d\mu \propto e^{f(x)} dx$ on the manifold, the corresponding Langevin dynamics to sample from μ can be written as the solution of a manifold-valued stochastic differential equation:

$$dX_t = \mathrm{grad} f(X_t) dt + \sqrt{2} dW_t$$

where grad denotes the Riemannian gradient, and W_t is Brownian motion on the manifold. See [LE20; CZS22] for more about this equation and discretization schemes, and [Hsu02] for a textbook treatment of stochastic calculus on manifolds.

For our results, all we need to know about this process is the form of the log-Sobolev inequality for the Langevin semigroup when the manifold is a product of n many $N-1$ dimensional spheres. As a special case of the general definition, we say the log-Sobolev inequality for the manifold Langevin dynamics for distribution μ holds with constant C if for all nonnegative functions f,

$$\mathrm{Ent}_\mu[f] \leq C \sum_{i=1}^n \mathbb{E}_\mu \left[\left\| \mathrm{grad}_i \sqrt{f} \right\|^2 \right], \tag{5}$$

where grad_i is the spherical gradient w.r.t. site i. This is the same inequality studied in the previous work [BB19] on the $O(N)$ model. Via the Herbst argument, it implies subgaussian concentration of Lipschitz functions. See [BGL+14; Hsu02] for more background on the log-Sobolev inequality and diffusion processes.

2.4 Stochastic localization

Given a measure ν_0 on $\mathbb{R}^n$, a standard n-dimensional Brownian motion W_t and an adapted process C_t valued in positive semidefinite $n \times n$ matrices, define the *Stochastic Localization* (SL) process to be the solution to the stochastic differential equation (SDE)

$$dF_t(x) = F_t(x) \cdot \langle x - \mathrm{mean}(\nu_t), C_t dW_t \rangle, \tag{6}$$

where $dv_t = F_t dv_0$ and $F_0 \equiv 1$. See, e.g., [Eld13; EMZ20; EKZ21] for the proof of the existence and uniqueness of this SDE for a sufficiently regular process C_t — in this work, we will only need the case where the driving matrix C_t is constant. The stochastic process v_t is a probability-measure-valued martingale, which means that $v_t(A)$ is a martingale for any measurable event A.

We will use the following Gaussian channel interpretation (see [EM22; KP21]) of stochastic localization with a time-invariant driving matrix.

Theorem 32 (Theorem 2 of [EM22])**.** *Let C be a positive definite $n \times n$ matrix and v_0 a distribution over $\mathbb{R}^n$. Let $X^* \sim v_0$, let B_t be an independent standard Brownian motion, and define*

$$y_t = tX^* + C^{-1}B_t.$$

Define $dv_t = F_t dv_0$ where the Radon-Nikodym derivative F_t is defined so that $\mathbb{E}_{v_0}[F_t] = 1$ and

$$F_t(x) \propto \exp\left(\langle C^2 y_t, x\rangle - \frac{t}{2}\langle x, C^2 x\rangle\right).$$

Then there exists a Brownian motion W_t adapted to the filtration generated by the y_t process such that F_t is a solution of the SDE (6) with $C_t = C$.

In this setting, the SL process is closely related to the Föllmer drift (see, e.g., [KP21]) and the Polchinski (renormalization group) equation, see [BD22; CE22]. We remark that in Ising models, if we look only at time $t = 1$ and choose driving matrix $C = J^{1/2}$, then y_1 is just the output of the Hubbard-Stratonovich transform.

Entropic stability. We also need to use some results concerning the key concept of *entropic stability* which we now recall. As a reminder, entropic stability itself arises as a generalization of the concept of entropic independence [Ana+21a], well-suited for simplicial localization schemes, to other linear-tilt localization schemes [CE22]. In particular, it tells us the exact analogue of entropic independence for stochastic localization.

Definition 33 (Exponential tilt)**.** Given a probability measure μ on $\mathbb{R}^n$, we let $\mathcal{T}_w\mu$ denote the *exponential tilt* by w. This is the probability measure with Radon-Nikodym derivative

$$\frac{d\mathcal{T}_w\mu}{d\mu}(x) \propto e^{\langle w, x\rangle}$$

provided that $\mathbb{E}_{x\sim\mu}\left[e^{\langle w,x\rangle}\right] < \infty$.

In the literature, an exponential tilt is sometimes called an *external field*.

Definition 34 (Entropic stability, Definition 29 of [CE22])**.** For μ, a probability measure on $\mathbb{R}^n$, a function $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, and $\alpha > 0$ we say that μ is *α-entropically stable* with respect to ψ if for every $w \in \mathbb{R}^n$, the exponential tilt $\mathcal{T}_w\mu$ exists and

$$\psi\big(\mathrm{mean}(\mathcal{T}_w\mu), \mathrm{mean}(\mu)\big) \leq \alpha\, \mathcal{D}_{\mathrm{KL}}(\mathcal{T}_w\mu \parallel \mu).$$

Entropic stability with respect to a quadratic functional induced by the driving matrix implies the following conservation of entropy estimate for stochastic localization.

Proposition 35 (Proposition 39 of [CE22])**.** *Let v_t be the stochastic localization process with driving matrix C_t, and suppose that for some $T > 0$ and almost surely for all $t \in [0, T]$ that v_t is α_t-entropically stable with respect to the function $\psi(x, y) = \frac{1}{2}\|C_t(x - y)\|_2^2$. Then for all nonnegative functions f,*

$$\mathbb{E}\left[\mathrm{Ent}_{v_T}[f]\right] \geq \exp\left(-\int_0^T \alpha_t dt\right) \cdot \mathrm{Ent}_{v_0}[f].$$

The following connects entropic stability with the covariance matrix of tilted measures.

Lemma 36 (Lemma 40 of [CE22])**.** *Let μ be a probability measure on $\mathbb{R}^n$, and let A and C be positive definite matrices. If $\mathrm{cov}(\mathcal{T}_w\mu) \preceq A$ for every $w \in \mathbb{R}^n$, then μ is α-entropically stable with respect to $\psi(x,y) = \frac{1}{2}\|C(x-y)\|_2^2$ for $\alpha = \|CAC\|_{\mathrm{op}}$.*

The property of having bounded covariance under all tilts is known as *semi-log-concavity* [ES22]; it is a weakening of the assumption of strong log-concavity of probability measures on $\mathbb{R}^n$ due to the Brascamp-Lieb inequality [BL02]. This property is equivalent to the assumption of the previous lemma after an appropriate change of basis, and weaker than fractional log-concavity of the generating polynomial defined in the next section.

Definition 37 (Semi-log-concavity)**.** Let μ be a probability measure on $\mathbb{R}^n$. We say that μ is β-semi-log-concave if $\mathrm{cov}(\mathcal{T}_w\mu) \preceq \beta I$ for every $w \in \mathbb{R}^n$.

Related to semi-log-concavity, defining the following semigroup on probability measures will simplify the notation in several of our arguments. It is related to the action of the heat semigroup on the Fourier transform, but it is nonlinear due to the presence of the normalization factor. Note that it commutes with the exponential tilt operation $\mathcal{T}_w$, and that it acts as the identity operator for measures whose support lies on a sphere of constant radius.

Definition 38. Given a probability measure μ on $\mathbb{R}^N$ and $s \in \mathbb{R}$, define probability measure $\mathcal{G}_s\mu$ by its relative density

$$\frac{d\mathcal{G}_s\mu}{d\mu}(x) \propto \exp(-s\|x\|^2/2)$$

assuming that the integral defining the normalizing constant exists (this is automatically true when $s \geq 0$).

Supermartingale property of relative entropy. Finally, we give an application of Lemma 29 in the context of stochastic localization. We will use this along with entropic stability to prove entropy contraction (i.e., approximate tensorization of entropy) estimates for down-up walks as well as the polarized walk. The very general form of the estimate is crucial for the latter application. For the Glauber dynamics, this was previously observed for the variance functional in [EKZ21], and a special case of this result with entropy was previously used by Lee [Lee23] to prove approximate tensorization of entropy estimates.

Lemma 39. *Let ν_t be the stochastic localization process with driving matrix C_t. For all Markov kernels κ and nonnegative functions f, the stochastic process (indexed by time $t \geq 0$)*

$$\mathbb{E}_{y\sim\nu_t\kappa}\left[\mathrm{Ent}_{\nu_t|\kappa\triangleright y}[f]\right]$$

is a supermartingale. Equivalently, for all $0 \leq s < t$ we have

$$\mathbb{E}\left[\mathbb{E}_{y\sim\nu_t\kappa}\left[\mathrm{Ent}_{\nu_t|\kappa\triangleright y}[f]\right] \,\middle|\, \mathcal{F}_s\right] \leq \mathbb{E}_{y\sim\nu_s\kappa}\left[\mathrm{Ent}_{\nu_s|\kappa\triangleright y}[f]\right],$$

where $\mathcal{F}_s$ is the filtration generated by the underlying Brownian motion in stochastic localization.

Proof. By the martingale property of stochastic localization, we have the mixture decomposition

$$\nu_s = \mathbb{E}[\nu_t \mid \mathcal{F}_s],$$

so the result follows immediately from Lemma 29. □

2.5 Geometry of polynomials

Definition 40 (Strongly log-concave polynomials [Gur09])**.** We say that a multivariate polynomial $f \in \mathbb{R}[z_1, \ldots, z_n]$ with nonnegative coefficients is strongly log-concave if for each of its derivatives $g = \partial_{i_1} \cdots \partial_{i_k} f$, either $g = 0$, or $\log g$ is concave over $\mathbb{R}^n_{>0}$.

When f is homogeneous, strong log-concavity is equivalent to complete log-concavity [Ana+18, Theorem 3.2], also known as Lorentzianity [BH20, Theorem 2.30]. See, e.g., [Ana+18; Ana+19a; Gur09; BH20] for more background.

We now recall a few key facts we will need about strongly log-concave polynomials. We specialize some facts to the homogeneous case since this is what we will use in this paper.

Proposition 41 (Proposition 2.7 of [Gur09], Proposition 5.1 of [Ana+18])**.** *A homogeneous bivariate polynomial $f(y, z) = \sum_{k=0}^{n} c_k y^{n-k} z^k$ is strongly log-concave iff its coefficients form an ultra log-concave sequence, that is $\{k \mid c_k > 0\}$ is a contiguous interval of integers, and for all $1 \le k \le n-1$,*

$$\left(\frac{c_k}{\binom{n}{k}}\right)^2 \ge \frac{c_{k-1}}{\binom{n}{k-1}} \cdot \frac{c_{k+1}}{\binom{n}{k+1}}.$$

Recall that the *elementary symmetric polynomial* e_k in variables $z_1, \ldots, z_n$ is

$$e_k(z) = \sum_{S \in \binom{[n]}{k}} \prod_{i \in S} z_i.$$

The elementary symmetric polynomials play a key role in *polarization*, which we define next. Polarization was first studied in the context of stable polynomials, i.e., polynomials that are root-free in a half-plane of the complex plane [BB08].

Definition 42 (Polarization)**.** Suppose that $g = \sum_{\alpha \in \mathbb{Z}^n_{\ge 0}} c_\alpha \prod_{i=1}^n z_i^{\alpha_i}$ is a polynomial that has degree at most κ_i in variable z_i. Define the polarization of g to be the multi-affine polynomial in variables (z_{ij}) for $i \in [n]$ and $j \in [\kappa_i]$ defined by

$$\mathrm{polar}_\kappa(g) = \sum_\alpha c_\alpha \prod_{i=1}^n \frac{e_{\alpha_i}(z_{i1}, \ldots, z_{i\kappa_i})}{\binom{\kappa_i}{a_i}}.$$

When κ is clear from context, we simply drop it and write polar for polarization.

Proposition 43 (Proposition 18 of [Ana+21c], [BH20])**.** *If a homogeneous polynomial g is strongly log-concave, then so is its polarization* $\mathrm{polar}_\kappa(g)$.

Strong log-concavity and similar properties of polynomials extend to distributions through the concept of generating polynomials.

Definition 44 (Generating polynomial)**.** If μ is a probability distribution on $2^{[n]}$, we define its generating polynomial to be

$$g_\mu(z_1, \ldots, z_n) = \mathbb{E}_{S \sim \mu}\left[\prod_{i \in S} z_i\right] = \sum_{S \subseteq [n]} \mu(S) \prod_{i \in S} z_i.$$

Next, we define a generalization of strong log-concavity.

Definition 45 (Fractional log-concavity)**.** For $\alpha \in (0,1]$, we say a multi-affine polynomial $g \in \mathbb{R}[z_1,\dots,z_n]$ with nonnegative coefficients is α-fractionally-log-concave (α-FLC) if $\log g(z_1^\alpha,\dots,z_n^\alpha)$ is concave on $\mathbb{R}^n_{>0}$.

We note that if a multi-affine g is α-FLC, so are its derivatives, since they can be derived as limits:

$$\partial_i g = \lim_{z_i \to \infty} g/z_i.$$

In particular, for multi-affine polynomials, 1-FLC is the same as strong log-concavity. We say a probability distribution on $2^{[n]}$ is α-FLC if its generating polynomial is α-FLC.

Definition 46 (Homogenization)**.** For a multi-affine polynomial $g \in \mathbb{R}[z_1,\dots,z_n]$, we define its multi-affine homogenization to be $g^{\hom} \in \mathbb{R}[z_1,\dots,z_n,\bar{z}_1,\dots,\bar{z}_n]$ defined as

$$g^{\hom}(z_1,\dots,z_n,\bar{z}_1,\dots,\bar{z}_n) = \bar{z}_1\cdots\bar{z}_n g\left(\frac{z_1}{\bar{z}_1},\dots,\frac{z_n}{\bar{z}_n}\right).$$

Similarly, for a distribution μ on $2^{[n]}$, we define $\mu^{\hom}$ to be the distribution whose generating polynomial is $g_\mu^{\hom}$. In other words, $\mu^{\hom}$ will be a distribution supported on $\binom{[n]\sqcup[\bar{n}]}{n} \subset 2^{[n]\sqcup[\bar{n}]}$, where for each $S \subseteq [n]$ we let

$$\mu^{\hom}\big(\{i \mid i \in S\} \sqcup \{\bar{i} \mid i \notin S\}\big) = \mu(S).$$

We will also often deal with distributions μ on $\{\pm 1\}^n$. Identifying $\{\pm1\}^n$ with $2^{[n]}$, we define the generating polynomial of such a μ as

$$g_\mu(z_1,\dots,z_n) = \mathbb{E}_{x\sim\mu}\left[\prod_{i:x_i=+1} z_i\right] = \sum_{x\in\{\pm1\}^n} \mu(x) \prod_{i:x_i=+1} z_i,$$

and its homogenization as

$$g_\mu^{\hom}(z_1,\dots,z_n,\bar{z}_1,\dots,\bar{z}_n) = \mathbb{E}_{x\sim\mu}\left[\prod_{i:x_i=+1} z_i \cdot \prod_{i:x_i=-1} \bar{z}_i\right] = \sum_x \mu(x) \prod_{i:x_i=+1} z_i \prod_{i:x_i=-1} \bar{z}_i.$$

We also associate another homogeneous multi-affine polynomial with μ by first homogenizing with a single additional variable, followed by polarization:

$$\begin{aligned}\operatorname{polar} g_\mu^{\hom}(z_1,\dots,z_n,y,\dots,y) &= \operatorname{polar}\mathbb{E}_{x\sim\mu}\left[y^{|\{i|x_i=-1\}|}\cdot\prod_{i:x_i=1} z_i\right] \\ &= \mathbb{E}_{x\sim\mu}\left[\frac{e_{|\{i|x_i=-1\}|}(y_1,\dots,y_n)}{\binom{n}{|\{i|x_i=-1\}|}}\cdot\prod_{i:x_i=1} z_i\right].\end{aligned}$$

We define the corresponding distribution as the *polar homogenization* of ν.

Definition 47 (Polar homogenization)**.** For a given n, let Π denote the Markov kernel from $2^{[n]}$ to $\binom{[n]\sqcup[\bar{n}]}{n}$, which maps $S \subseteq [n]$ to $S \sqcup T$, where $T \in \binom{[\bar{n}]}{n-|S|}$ is chosen uniformly at random. For a distribution μ on $2^{[n]}$ we call $\mu\Pi$ the polar homogenization of μ. This is a distribution supported on sets $S \sqcup T$ where $S \subseteq [n]$ and $T \subseteq [\bar{n}]$ and $|S| + |T| = n$:

$$\mu\Pi(S \sqcup T) = \frac{\mu(S)}{\binom{n}{n-|S|}} = \frac{\mu(S)}{\binom{n}{|S|}}.$$

We extend the definition of exponential tilts, Definition 33, to distributions μ on $2^{[n]}$ by identifying $2^{[n]}$ with $\{0,1\}^n \subset \mathbb{R}^n$. In other words,

$$\mathcal{T}_w \mu(S) = \frac{\mu(S) \prod_{i \in S} e^{w_i}}{g_\mu(e^{w_1}, \ldots, e^{w_n})}$$

where g_μ, the generating polynomial of μ, gives us the normalizing constant. Note that α-fractional-log-concavity is closed under exponential tilts.

Definition 48 (Correlation matrix)**.** Let μ be a probability distribution over $2^{[n]}$. Its correlation matrix $\Psi^{\mathrm{cor}}_\mu \in \mathbb{R}^{n \times n}$ is defined by

$$\Psi^{\mathrm{cor}}_\mu(i,j) = \mathbb{P}_{S \sim \mu}[j \in S \mid i \in S] - \mathbb{P}_{S \sim \mu}[i \in S].$$

Definition 49 (Spectral independence)**.** For $\eta \geq 0$, a distribution μ on $2^{[n]}$ is said to be η-spectrally independent if

$$\lambda_{\max}(\Psi^{\mathrm{cor}}_\mu) \leq \eta.$$

Remark 50. The original definition of spectral independence in [ALO20] imposes this requirement on μ as well as all of its "links." Here, we follow the convention in [Ana+21a] and use the term spectral independence to refer only to a spectral norm bound on the correlation matrix of μ.

Proposition 51 (Remark 70 of [Ali+21])**.** *A distribution μ on $2^{[n]}$ is η-spectrally independent iff*

$$\nabla^2 \log g_\mu\left(z_1^{1/\eta}, \ldots, z_n^{1/\eta}\right)\Big|_{z=\mathbb{1}} = (1/\eta)^2 D\Psi^{\mathrm{cor}}_\mu - (1/\eta)D \preceq 0,$$

where D is the diagonal matrix with entries $D_{ii} = \mathbb{P}_{S \sim \mu}[i \in S]$. Moreover, μ is $1/\eta$-FLC iff $\mathcal{T}_w \mu$ is η-spectrally independent for all $w \in \mathbb{R}^n$.

With some abuse of notation, for distributions μ on $2^{[n]}$ we use $\mathrm{cov}(\mu)$ and $\mathrm{mean}(\mu)$ to denote $\mathrm{cov}(\mathbb{1}_S)$ and $\mathrm{mean}(\mathbb{1}_S)$ where $S \sim \mu$. We note that $D\Psi^{\mathrm{cor}}_\mu$ is the same as the $\mathrm{cov}(\mu)$.

Spectral independence is frequently used to obtain Var contraction rates for *down kernels*.

Definition 52. For a fixed ground set $[n]$ and $0 \leq \ell \leq k \leq n$, we let $D_{k \to \ell}$ denote the Markov kernel from $\binom{[n]}{k}$ to $\binom{[n]}{\ell}$ defined as follows: given a set S, we map it to a uniformly random subset $T \in \binom{S}{\ell}$.

For a distribution μ on $\binom{[n]}{k}$, we have η/k contraction of Var under the kernel $D_{k \to 1}$ iff μ is η-spectrally independent [Ali+21, see, e.g.,].

3 Trickle-down equation in linear-tilt localization schemes

3.1 Trickle-down equation

Linear-tilt localization scheme. Let ν_0 be a probability measure. Suppose there exists a sequence of random vectors Z_t generating filtrations $\mathcal{F}_t = \sigma(Z_0, \ldots, Z_t)$ so that $\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t] = 0$, i.e., they form a martingale difference sequence. Define the corresponding measure-valued stochastic process $(\nu_t)_t$ by its Radon-Nikodym derivative

$$\frac{d\nu_{t+1}}{d\nu_t}(x) = 1 + \langle x - \mathrm{mean}(\nu_t), Z_{t+1} \rangle.$$

Note that

$$\mathbb{E}_{\nu_t}\left[\frac{d\nu_{t+1}}{d\nu_t}\right] = 1 + \langle \mathrm{mean}(\nu_t) - \mathrm{mean}(\nu_t), Z_{t+1} \rangle = 1,$$

so, provided we check nonnegativity, ν_{t+1} is indeed a probability measure (almost surely). To ensure nonnegativity and guarantee ν_{t+1} is a probability measure, we make the assumption

$$\mathbb{P}_{x\sim\nu_t}[\langle x - \mathrm{mean}(\nu_t), Z_{t+1}\rangle \geq -1] = 1. \tag{7}$$

Also, observe that for any measurable set S

$$\mathbb{E}[\nu_{t+1}(S) \mid \mathcal{F}_t] = \nu_t(S),$$

because $\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t] = 0$. So $(\nu_t)_t$ is a measure-valued martingale.

We say this process is a *linear-tilt localization scheme* if additionally, for any measurable event S, $\nu_t(S)$ almost surely converges to either 0 or 1 as $t \to \infty$. This is consistent with the terminology of Section 2.4 of [CE22]. However, the calculations we perform next do not require this additional assumption.

Trickle-down equation. Now we formulate a general trickle-down recursion. Recall that we have the following *law of total variance* for ν_t:

$$\begin{aligned}
\mathrm{cov}(\nu_t) &= \mathbb{E}_{x\sim\nu_t}[x^{\otimes 2}] - \mathbb{E}_{x\sim\nu_t}[x]^{\otimes 2} = \mathbb{E}_{x\sim\nu_t}[x^{\otimes 2}] - \mathrm{mean}(\nu_t)^{\otimes 2} \\
&= \mathbb{E}\left[\mathbb{E}_{x\sim\nu_{t+1}}[x^{\otimes 2}] - \mathrm{mean}(\nu_{t+1})^{\otimes 2} \;\middle|\; \mathcal{F}_t\right] + \mathbb{E}\left[\mathrm{mean}(\nu_{t+1})^{\otimes 2} - \mathrm{mean}(\nu_t)^{\otimes 2} \;\middle|\; \mathcal{F}_t\right] \\
&= \mathbb{E}[\mathrm{cov}(\nu_{t+1}) \mid \mathcal{F}_t] + \mathrm{cov}(\mathrm{mean}(\nu_{t+1}) \mid \mathcal{F}_t).
\end{aligned}$$

Observe that by definition

$$\begin{aligned}
\mathrm{mean}(\nu_{t+1}) - \mathrm{mean}(\nu_t) &= \mathbb{E}_{x\sim\nu_t}[x \cdot \langle x - \mathrm{mean}(\nu_t), Z_{t+1}\rangle] \\
&= \mathbb{E}_{x\sim\nu_t}[x \otimes (x - \mathrm{mean}(\nu_t))]Z_{t+1} = \mathrm{cov}(\nu_t)Z_{t+1}
\end{aligned}$$

so

$$\begin{aligned}
\mathrm{cov}(\nu_t) &= \mathbb{E}[\mathrm{cov}(\nu_{t+1}) \mid \mathcal{F}_t] + \mathbb{E}\left[(\mathrm{mean}(\nu_{t+1}) - \mathrm{mean}(\nu_t))^{\otimes 2} \;\middle|\; \mathcal{F}_t\right] \\
&= \mathbb{E}[\mathrm{cov}(\nu_{t+1}) \mid \mathcal{F}_t] + \mathrm{cov}(\nu_t)\,\mathbb{E}\left[Z_{t+1}^{\otimes 2} \;\middle|\; \mathcal{F}_t\right]\mathrm{cov}(\nu_t) \\
&= \mathbb{E}[\mathrm{cov}(\nu_{t+1}) \mid \mathcal{F}_t] + \mathrm{cov}(\nu_t)\,\mathrm{cov}(Z_{t+1} \mid \mathcal{F}_t)\,\mathrm{cov}(\nu_t)
\end{aligned} \tag{8}$$

or rearranging, we have

$$\mathrm{cov}(\nu_t) - \mathrm{cov}(\nu_t)\,\mathrm{cov}(Z_{t+1} \mid \mathcal{F}_t)\,\mathrm{cov}(\nu_t) = \mathbb{E}[\mathrm{cov}(\nu_{t+1}) \mid \mathcal{F}_t].$$

We can recursively apply Eq. (8) up to any time $T \geq t$ to yield the more general integral formula

$$\mathrm{cov}(\nu_t) = \mathbb{E}[\mathrm{cov}(\nu_T) \mid \mathcal{F}_t] + \sum_{s=t}^{T-1} \mathbb{E}[\mathrm{cov}(\nu_s)\,\mathrm{cov}(Z_{s+1} \mid \mathcal{F}_s)\,\mathrm{cov}(\nu_s) \mid \mathcal{F}_t] \tag{9}$$

which will be very useful in the present work.

3.2 Examples

Example: stochastic localization. Recall first that stochastic localization is the SDE with driving matrix C_t defined as

$$dF_t(x) = F_t(x)\langle x - \mathrm{mean}(\nu_t), C_t\,dW_t\rangle$$

where W_t is a brownian motion and $d\nu_t = F_t\,d\nu_0$.

For $\Delta > 0$, which we will take to zero, we can parameterize time in multiples of Δ. The Euler-Murayama discretization of the SDE for stochastic localization with driving matrix C_t would be

$$\frac{d\nu_{t+\Delta}}{d\nu_t}(x) = 1 + \langle x - \mathrm{mean}(\nu_t), Z_{t+\Delta}\rangle$$

where $Z_{t+\Delta} \sim \mathcal{N}(0, \Delta C_t^2)$. Taking the limit of $\Delta \to 0$, Eq. (9) becomes

$$\mathrm{cov}(\nu_t) = \mathbb{E}[\mathrm{cov}(\nu_T) \mid \mathcal{F}_t] + \int_t^T \mathbb{E}[\mathrm{cov}(\nu_s) C_s^2 \,\mathrm{cov}(\nu_s) \mid \mathcal{F}_t] ds \tag{10}$$

which is the trickle-down equation for stochastic localization. To be precise, the discretized process with fixed $\Delta > 0$ is not a localization scheme since it can assign events negative probabilities; however, it is a valid localization scheme in the limit $\Delta \to 0$. For completeness, we include a more formal proof of Eq. (9) below.

Theorem 53. *Let ν_0 be a probability measure on $\mathbb{R}^n$ and let $(\nu_t)_{t\geq 0}$ be the stochastic localization process with driving matrix C_t. Then Eq. (10) holds for any $0 \leq t \leq T$.*

Proof. This is the same proof as above, except executed rigorously in the limit $\Delta \to 0$ using stochastic calculus. It is also a standard calculation in the literature on stochastic localization, see e.g. [Eld20]. We have

$$d\,\mathrm{mean}(\nu_t) = \mathbb{E}_{x\sim\nu_t}[x \cdot \langle x - \mathrm{mean}(\nu_t), C_t dW_t\rangle] = \mathbb{E}_{x\sim\nu_t}[x \otimes (x - \mathrm{mean}(\nu_t))] C_t dW_t = \mathrm{cov}(\nu_t) C_t dW_t.$$

Applying the Itô formula, we get

$$\begin{aligned} d(\mathrm{mean}(\nu_t)^{\otimes 2}) &= \mathrm{cov}(\nu_t) C_t^2 \,\mathrm{cov}(\nu_t) dt + \mathrm{mean}(\nu_t) \otimes (d\,\mathrm{mean}(\nu_t)) + (d\,\mathrm{mean}(\nu_t)) \otimes \mathrm{mean}(\nu_t) \\ &= \mathrm{cov}(\nu_t) C_t^2 \,\mathrm{cov}(\nu_t) dt + \text{martingale}, \end{aligned}$$

where martingale denotes a pure diffusion term, i.e., one with no drift. Since $\mathrm{cov}(\nu_t) = \mathbb{E}_{x\sim\nu_t}[x^{\otimes 2}] - \mathrm{mean}(\nu_t)^{\otimes 2}$, and $\mathbb{E}_{x\sim\nu_t}[x^{\otimes 2}]$ is a martingale, we have

$$d\,\mathrm{cov}(\nu_t) = d\,\mathbb{E}_{x\sim\nu_t}[x^{\otimes 2}] - d(\mathrm{mean}(\nu_t)^{\otimes 2}) = \text{martingale} - \mathrm{cov}(\nu_t) C_t^2 \,\mathrm{cov}(\nu_t) dt.$$

Integrating and taking the conditional expectation, which cancels the martingale terms, it follows that for any $T \geq t$,

$$\mathrm{cov}(\nu_t) = \mathbb{E}[\mathrm{cov}(\nu_T) \mid \mathcal{F}_t] + \int_t^T \mathbb{E}[\mathrm{cov}(\nu_s) C_s^2 \,\mathrm{cov}(\nu_s) \mid \mathcal{F}_t] ds$$

as claimed. □

Example: pinning in set systems (pure simplicial complex). Suppose that ν_0 is a probability measure on $\binom{[n]}{k}$ which we embed in $\{0,1\}^n \subset \mathbb{R}^n$ by mapping $S \in \binom{[n]}{k}$ to $\mathbb{1}_s$. Following the high-dimensional expanders paradigm, we can naturally define a sequence of measures in the following way: let $S \sim \nu_0$, let $(a_1, \ldots, a_k)$ be the elements of S ordered according to a uniformly random permutation, and for $t \in \{0, \ldots, k\}$ define

$$\nu_t = \nu_0 \mid D_{k\to t} \triangleright \{a_1, \ldots, a_t\},$$

where $D_{k\to t}$ is the down kernel from Definition 52. In other words, ν_t is the posterior distribution of $T \sim \nu_0$ given the observation $\{a_1, \ldots, a_t\} \subseteq T$. We can optionally define $\nu_t = \nu_k$ for every $t > k$ to

be consistent with the definition of localization schemes. Also, it is worth noting that the "link" in high-dimensional expanders terminology will not be ν_t, but the law of $S \setminus \{a_1, \dots, a_t\}$ for $S \sim \nu_t$.

This is a linear-tilt localization scheme[5] w.r.t. the filtration $\mathcal{F}_t = \sigma(\{a_1, \dots, a_t\})$. This is because

$$\begin{aligned}
\nu_{t+1}(T) &= \frac{\nu_t(T) \cdot \mathbb{1}[a_{t+1} \in T]}{\mathbb{P}_{R\sim\nu_t}[a_{t+1} \in R]} = \nu_t(T) \cdot \left(1 + \frac{\mathbb{1}[a_{t+1} \in T] - \mathbb{P}_{R\sim\nu_t}[a_{t+1} \in R]}{\mathbb{P}_{R\sim\nu_t}[a_{t+1} \in R]}\right) \\
&= \nu_t(T) \cdot \left(1 + \left\langle \mathbb{1}_T - \mathrm{mean}(\nu_t), \frac{\mathbb{1}_{a_{t+1}}}{\mathbb{P}_{R\sim\nu_t}[a_{t+1} \in R]} \right\rangle\right) \\
&= \nu_t(T) \cdot (1 + \langle \mathbb{1}_T - \mathrm{mean}(\nu_t), Z_{t+1} \rangle)
\end{aligned}$$

where we define $Y_{t+1} = \mathbb{1}_{a_{t+1}}/\mathbb{P}_{R\sim\nu_t}[a_{t+1} \in R]$ and $Z_{t+1} = Y_{t+1} - \mathbb{E}[Y_{t+1} \mid \mathcal{F}_t]$. Here we used $\langle \mathbb{1}_T - \mathrm{mean}(\nu_t), \mathbb{E}[Y_{t+1} \mid \mathcal{F}_t] \rangle = 0$, which follows from Eq. (11) below and observing that for any R in the support of ν_t we have $\langle \mathbb{1}_R, \mathbb{1}_{[n]\setminus\{a_1,\dots,a_t\}} \rangle = k - t$, which also means $\langle \mathrm{mean}(\nu_t), \mathbb{1}_{[n]\setminus\{a_1,\dots,a_t\}} \rangle = k - t$.

From the definition of the random vectors Z_t, it is clear that they generate the same filtration $\mathcal{F}_t$ as the variables a_t and that they form a martingale difference sequence. We observe that the conditional law of a_{t+1} given $\mathcal{F}_t$ is the same as picking an element uniformly at random from $T \setminus \{a_1, \dots, a_t\}$ where $T \sim \nu_t$. Using this we have

$$\mathbb{E}[Y_{t+1} \mid \mathcal{F}_t] = \sum_{e \in [n]\setminus a_1,\dots,a_t} \frac{\mathbb{P}_{R\sim\nu_t}[e \in R]}{k-t} \cdot \frac{\mathbb{1}_e}{\mathbb{P}_{R\sim\nu_t}[e \in R]} = \frac{\mathbb{1}_{[n]\setminus\{a_1,\dots,a_t\}}}{k-t}. \tag{11}$$

So

$$\begin{aligned}
\mathrm{cov}(Z_{t+1} \mid \mathcal{F}_t) &= \mathbb{E}\left[Y_{t+1}^{\otimes 2} \,\middle|\, \mathcal{F}_t\right] - \mathbb{E}[Y_{t+1} \mid F_t]^{\otimes 2} \\
&= \frac{1}{k-t} \cdot \left(\sum_{e \in [n]\setminus\{a_1,\dots,a_t\}} \frac{\mathbb{1}_e^{\otimes 2}}{\mathbb{P}_{R\sim\nu_t}[e \in R]} \right) - \left(\frac{\mathbb{1}_{[n]\setminus\{a_1,\dots,a_t\}}}{k-t} \right)^{\otimes 2}.
\end{aligned}$$

Again, the vector $\mathbb{1}_{[n]\setminus\{a_1,\dots,a_t\}}$ is orthogonal to all $\mathbb{1}_T - \mathrm{mean}(\nu_t)$, and thus is in the kernel of $\mathrm{cov}(\nu_t)$. So Eq. (9) simplifies a bit, and we obtain the equation

$$\mathrm{cov}(\nu_t) = \mathbb{E}[\mathrm{cov}(\nu_T) \mid \mathcal{F}_t] + \sum_{s=t}^{T-1} \frac{1}{k-s} \mathbb{E}[\mathrm{cov}(\nu_s) N_s^{-1} \mathrm{cov}(\nu_s) \mid \mathcal{F}_t] \tag{12}$$

where N_s is a diagonal matrix encoding the marginals of the link at $\{a_1, \dots, a_s\}$; in other words, $N_s = \mathrm{diag}(\mathrm{mean}(\nu_s) - \mathbb{1}_{\{a_1,\dots,a_s\}})$. When we set $T = t + 1$, this is a familiar equation from which the trickle-down equation of Oppenheim [Opp18] and matrix trickle-down [ALG22] can both be derived. For the reader's convenience, we illustrate how this works for Oppenheim's trickle-down in Appendix A.1. We also give another example in a similar spirit in Appendix A.2 illustrating trickle-down of semi-log-concavity.

4 Rapid mixing of Glauber dynamics in Ising models

In this section, we prove the main result about rapid mixing of Ising models. The supporting lemmas are given after the proof of the main theorem. Some elements of the analysis can be done in a more general setting, but to keep the presentation concrete we hold off on making most generalizations until the next section of the paper.

[5] This observation generalizes one from Section 2.4.1 of [CE22] made in the context of spin systems.

4.1 Covariance bound and approximate tensorization

Theorem 54. *Suppose that ν is a probability measure on $\{\pm 1\}^n$ defined by*

$$\nu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle\right)$$

for some external field $h \in \mathbb{R}^n$ and interaction matrix J satisfying $J \succeq 0$. Without loss of generality, suppose that the diagonal of J is constant, so there exists α such that $J_{ii} = \alpha \geq 0$. Let $\eta = \alpha/\|J\|_{\mathrm{op}} \in [0,1]$. Then

$$\|\mathrm{cov}(\nu)\|_{\mathrm{op}} \leq q_\eta(\|J\|_{\mathrm{op}})$$

where $q_\eta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ solves the Volterra integral equation

$$q_\eta(z) = \big(1 - r(\eta z)\big) + \int_0^z q(y)^2 dy$$

where

$$r(a) = \mathbb{E}_{\sigma \sim \mathrm{Ber}_\pm(\tanh(0)), g \sim \mathcal{N}(0,1)}\left[\tanh^2(a\sigma + \sqrt{a}g)\right].$$

Furthermore, ν satisfies Approximate Tensorization of Entropy with constant at most

$$\exp\left(\int_0^{\|J\|_{\mathrm{op}}} q_\eta(z)dz\right).$$

Proof. Let ν_t be the stochastic localization process with driving matrix $C_t = J^{1/2}$ and $\nu_0 = \nu$, and define $\Sigma_t = \mathrm{cov}(\nu_t)$. Recall from Eq. (10) that

$$\Sigma_t = \mathbb{E}[\Sigma_1 \mid \mathcal{F}_t] + \int_t^1 \mathbb{E}[\Sigma_s J \Sigma_s \mid \mathcal{F}_t] ds$$

so applying Lemma 57 we have

$$\Sigma_t \preceq \mathrm{diag}(1 - r(\alpha(1-t))) + \int_t^1 \mathbb{E}[\Sigma_s J \Sigma_s \mid \mathcal{F}_t] ds.$$

Define $\beta(t) = \sup_h \|\mathrm{cov}(\mu_{(1-t)J,h})\|_{\mathrm{op}}$, where $\mu_{(1-t)J,h}$ denotes an Ising model with interaction matrix $(1-t)J$ and external field h. Then, by considering the stochastic localization process started from this model, we have that $\beta(1) = 1$ and

$$\beta(t) \leq [1 - r(\alpha(1-t))] + \|J\|_{\mathrm{op}} \int_t^1 \beta(s)^2 ds$$

so we can apply Lemma 55. Therefore

$$\|\Sigma_0\|_{\mathrm{op}} \leq \gamma(1)$$

where $\gamma(t)$ is the solution to (13) with $K = \|J\|_{\mathrm{op}}$ and $f(t) = 1 - r(\alpha t)$, so

$$\gamma(t) = [1 - r(\alpha t)] + \|J\|_{\mathrm{op}} \int_0^t \gamma(s)^2 ds.$$

Making the change of variables $q(z) = \gamma(z/\|J\|_{\mathrm{op}})$ (i.e., $z = \|J\|_{\mathrm{op}} t$) yields the integral equation

$$q_\eta(z) = [1 - r(\alpha z/\|J\|_{\mathrm{op}})] + \|J\|_{\mathrm{op}} \int_0^{z/\|J\|_{\mathrm{op}}} \gamma(s)^2 ds = [1 - r(\alpha z/\|J\|_{\mathrm{op}})] + \int_0^z q_\eta(y)^2 ds$$

where we made the corresponding change of variables $y = \|J\|_{\mathrm{op}} s$ inside the integral. Hence

$$\|\Sigma_0\|_{\mathrm{op}} \le q_\eta(\|J\|_{\mathrm{op}})$$

which proves the bound on the covariance matrix.

By Lemma 36, this implies that every such ν is $\|J\|_{\mathrm{op}} q_\eta(\|J\|_{\mathrm{op}})$-entropically stable with respect to the function $\psi(x, y) = \frac{1}{2}\|J^{1/2}(x - y)\|_2^2$. The measure ν_1 is a product measure and therefore satisfies ATE with constant 1. Therefore, for any nonnegative function f

$$\begin{aligned}
\mathrm{Ent}_{\nu_0}[f] &\le \exp\left(\int_0^1 q_\eta(t\|J\|_{\mathrm{op}})\|J\|_{\mathrm{op}} dt\right) \mathbb{E}[\mathrm{Ent}_{\nu_1}[f]] \\
&\le \exp\left(\int_0^1 q_\eta(t\|J\|_{\mathrm{op}})\|J\|_{\mathrm{op}} dt\right) \sum_i \mathbb{E}[\mathbb{E}_{X_{\sim i} \sim \nu_1}[\mathrm{Ent}_{\nu_1(\cdot|X_{\sim i})}[f]]] \\
&\le \exp\left(\int_0^1 q_\eta(t\|J\|_{\mathrm{op}})\|J\|_{\mathrm{op}} dt\right) \sum_i \mathbb{E}_{X_{\sim i} \sim \nu_0}[\mathrm{Ent}_{\nu_0(\cdot|X_{\sim i})}[f]]
\end{aligned}$$

where in the first inequality we used Proposition 35, the second inequality is tensorization for the product measure, and the last inequality is the supermartingale property (Lemma 39). Making the change of variables $z = t\|J\|_{\mathrm{op}}$ proves the result.

□

Lemma 55. *Suppose $f(t) \ge 0$ for $t \ge 0$ and $K \ge 0$. Either there exists a unique solution for all time, or there exists $T > 0$ and a solution on $[0, T)$ to the integral equation*

$$\gamma(t) = f(t) + K \int_0^t \gamma(s)^2 ds, \tag{13}$$

such that the solution is unique and satisfies $\gamma(t) \to \infty$ as $t \to T$. Suppose $\alpha(t)$ is a function satisfying

$$\alpha(t) \le f(t) + K \int_0^t \alpha(s)^2 ds.$$

Then $\alpha(t) \le \gamma(t)$.

Proof. See Chapter 5 of [LL69] — Theorem 5.3.1 establishes the comparison inequality, Theorem 5.2.1 establishes existence, and Theorem 5.4.4 establishes uniqueness. □

We recall the following basic fact.

Lemma 56. *Suppose that μ is a probability measure on $\{\pm 1\}$ such that $\mu(x) \propto \exp(hx)$. Then $\mathbb{E}_\mu[X] = \tanh(h)$ and $\mathrm{Var}_\mu[X] = 1 - \tanh^2(h)$.*

Proof. This follows from the definitions using that $\tanh(h) = \frac{e^h - e^{-h}}{e^h + e^{-h}}$. □

Lemma 57. *Under the assumptions of Theorem 54, let ν_t be the stochastic localization process with driving matrix $C_t = J^{1/2}$. Then for any $t \in [0,1]$,*

$$\mathbb{E}[\Sigma_1 \mid \mathcal{F}_t] \preceq \mathrm{diag}([1 - r(\alpha \cdot (1-t))]_{i \in [n]})$$

where r is defined in Lemma 58 and $\Sigma_1 = \mathrm{cov}(\nu_1)$.

Proof. We first prove the result when $t = 0$ and then generalize. Recall from Theorem 32 that the measure ν_1 is equal in law to the product measure

$$\mu(x) \propto \exp(\langle JY_1 + h, x \rangle)$$

where $Y_1 = X^* + J^{-1/2}G$ is generated by sampling $X^* \sim \nu_0$ and independently $G \sim \mathcal{N}(0, I)$, so that

$$JY_1 + h = JX^* + h + J^{1/2}G.$$

Hence Σ_1 is diagonal and furthermore by applying Lemma 56 we have

$$\mathbb{E}[(\Sigma_1)_{ii}] = 1 - \mathbb{E}_{X^* \sim \nu_0, g \sim \mathcal{N}(0,1)}[\tanh^2(\langle J_i, X^* \rangle + h_i + \sqrt{J_{ii}} g)] \leq 1 - r(J_{ii})$$

where r is defined in Lemma 58 which we applied to get the final bound. This proves the result in the special case $t = 0$.

We now consider the case where $t > 0$. We know, either by explicit calculation or by Theorem 32, that ν_t is an Ising model with interaction matrix $(1-t)J$ and external field h_t. Furthermore, conditional on $\mathcal{F}_t$ the measure ν_1 is equal in law to a product measure

$$\mu(x) \propto \exp(\langle JY_1' + h_t, x \rangle)$$

where $X' \sim \nu_t$

$$Y_1' = (1-t)X^* + \sqrt{1-t}\, J^{-1/2}G'$$

with $G' \sim \mathcal{N}(0, I)$ so that

$$(1-t)JY_1' + h_t = (1-t)JX' + h_t + \sqrt{1-t} J^{1/2} G.$$

Hence applying Lemma 58 in the same way proves the result in general. □

Lemma 58. *Suppose that ν is an Ising model on n sites with interaction matrix J and external field h. Then for any $i \in [n]$*

$$\mathbb{E}_{X^* \sim \nu_0, g \sim \mathcal{N}(0,1)}[\tanh^2(\langle J_i, X^* \rangle + h_i + \sqrt{J_{ii}} g)] \geq \inf_b r(J_{ii})$$

where

$$r(a) = \mathbb{E}_{\sigma \sim \mathrm{Ber}_{\pm}(0), g \sim \mathcal{N}(0,1)}[\tanh^2(a\sigma + \sqrt{a} g)]$$

and $\mathrm{Ber}_{\pm}(z)$ is the law of a random variable valued in $\{\pm 1\}$ with mean z.

Proof. Using that

$$\mathbb{E}[X_i^* \mid X_{\sim i}^*] = \tanh(J_{i,\sim i} \cdot X_{\sim i}^* + h_i)$$

and considering $b = \langle J_{i,\sim i}, X_{\sim i}^* \rangle + h_i$, we can lower bound

$$\begin{aligned}\mathbb{E}_{X^* \sim \nu_0, g \sim \mathcal{N}(0,1)}[\tanh^2(\langle J_i, X^* \rangle + h_i + \sqrt{J_{ii}} g)] &= \mathbb{E}_{X^* \sim \nu_0, g \sim \mathcal{N}(0,1)}[\mathbb{E}[\tanh^2(\langle J_i, X^* \rangle + h_i + \sqrt{J_{ii}} g) \mid b]] \\ &\geq \inf_{b \in \mathbb{R}} \mathbb{E}_{\sigma \sim \mathrm{Ber}_{\pm}(\tanh(b)), g \sim \mathcal{N}(0,1)}[\tanh^2(b + J_{ii}\sigma + \sqrt{J_{ii}} g)].\end{aligned}$$

By Lemma 59, the infimum on the right hand side is attained at $b = 0$. □

4.2 A monotonicity inequality via coupling stochastic localizations

In this section we prove the following lemma as a consequence of a more general construction of a coupling of two stochastic localization processes, which we elaborate upon below.

Lemma 59. *For any $a \geq 0$, the global minimum of the function $t_a : \mathbb{R} \to \mathbb{R}$ defined by*

$$t_a(b) = \mathbb{E}_{\sigma \sim \mathrm{Ber}_{\pm}(\tanh(b)), g \sim \mathcal{N}(0,1)}[\tanh^2(b + a\sigma + \sqrt{a}g)]$$

is attained at $b = 0$.

Proof. It is equivalent to show that the global maximum of $1 - t_a$ is attained at $b = 0$. Using the Gaussian channel interpretation (Theorem 32), $1 - t_a(b)$ can be interpreted as the operator norm of the covariance matrix of stochastic localization initialized at the measure $\mathrm{Ber}_{\pm}(\tanh(b))$ and run for time a. Since $\tanh^2$ is monotonically increasing on $\mathbb{R}_{\geq 0}$ and symmetric, the assumption of Lemma 61 is satisfied, and so the desired conclusion follows. □

It remains to establish the desired fact about stochastic localization. Note that the stochastic localization process used here is different from the one in the previous section (the process here is operating on a single "spin" of the system, without interacting with the other spins). Informally, the general lemma established below allows us to lift certain monotonicity inequalities from the final measure arrived at in stochastic localization back to the initial measure.

The argument takes advantage of symmetry in a key way.

Definition 60. We say a probability measure μ on $\mathbb{R}^N$ is spherically symmetric if for every orthogonal matrix R and measurable set S, $\mu(RS) = \mu(S)$.

For example, when $N = 1$ spherical symmetry means that the measure is unchanged under the reflection map $x \mapsto -x$. We can now state a more general monotonicity result.

Lemma 61. *Suppose that μ is a probability measure in $\mathbb{R}^N$ which is spherically symmetric. Let $T \geq 0$. Suppose that the following reverse monotonicity inequality holds: for any $v, w \in \mathbb{R}^N$ such that $\|v\| \leq \|w\|$,*

$$\|\operatorname{cov}(\mathcal{T}_v \mathcal{G}_T \mu)\|_{OP} \geq \|\operatorname{cov}(\mathcal{T}_w \mathcal{G}_T \mu)\|_{OP}.$$

Then the following reverse monotonicity inequality holds. Let $v, w \in \mathbb{R}^N$ be such that $\|v\| \leq \|w\|$. Suppose that for $h \in \{v, w\}$, the law of stochastic process $\mu_t(h)$ for $t \geq 0$ is stochastic localization with identity driving matrix initialized at $\mu_0(h) = \mathcal{T}_h \mu$, i.e. initialized at

$$\frac{d\mu_0(h)}{d\mu}(x) \propto e^{\langle h, x \rangle}.$$

Then

$$\mathbb{E}[\|\operatorname{cov}(\mu_T(v))\|_{OP}] \geq \mathbb{E}[\|\operatorname{cov}(\mu_T(w))\|_{OP}].$$

Proof. This follows directly from Lemma 62, since it constructs a coupling where $\|\operatorname{cov}(\mu_T(v))\|_{OP} \geq \|\operatorname{cov}(\mu_T(w))\|_{OP}$ holds almost surely (in other words, one stochastically dominates the other). Taking expectation on both sides proves the desired result. □

Now we state and prove the crucial coupling result.

Lemma 62. *Suppose that μ is a probability measure in $\mathbb{R}^N$ which is rotationally symmetric. For any v, w with $\|v\| \leq \|w\|$ there exists a joint coupling of stochastic processes $\mu_t(v), \mu_t(w)$ indexed by $t \geq 0$ such that:*

1. *The marginal law of the stochastic process $(\mu_t(v))_t$ is stochastic localization initialized at $\mu_0(v) = \mathcal{T}_v\mu$ with driving matrix $C_t = I$.*
2. *Likewise, the marginal law of $(\mu_t(w))_t$ is stochastic localization initialized at $\mu_0(w) = \mathcal{T}_w\mu$ with driving matrix $C_t = I$.*
3. *There exist adapted $\mathbb{R}^N$-valued stochastic processes v_t, w_t such that $\mu_t(v) = \mathcal{T}_{v_t}\mathcal{G}_t\mu$, $\mu_t(w) = \mathcal{T}_{w_t}\mathcal{G}_t\mu$, and $\|v_t\| \le \|w_t\|$ for all times t almost surely.*

Proof. For simplicity, we give the proof below in the case that μ is a measure on a discrete set, but the same argument works in general with only minor changes.

The construction is a multi-step process. First we construct $\mu_t(v)$ for all times as stochastic localization driven by an N-dimensional Brownian motion W_t. Recall that the definition of stochastic localization driven by Brownian motion W_t is that

$$d\mu_t(v)(x) = \mu_t(v)(x)\langle x - \mathrm{mean}(\mu_t(v)), dW_t\rangle$$

so that by Ito's formula

$$\begin{aligned} d\log\mu_t(v)(x) &= \langle x - \mathrm{mean}(\mu_t(v)), dW_t\rangle - \frac{1}{2}\|x - \mathrm{mean}(\mu_t(v))\|^2 dt \\ &= \langle x, dW_t\rangle + \langle x, \mathrm{mean}(\mu_t(v))\rangle dt - \frac{1}{2}\|x\|^2 dt - \frac{1}{2}\|\mathrm{mean}(\mu_t(v))\|^2 dt - \langle \mathrm{mean}(\mu_t(v)), dW_t\rangle. \end{aligned}$$

where we note that the last two terms are independent of x. Hence if we define v_t by $v_0 = v$ and $dv_t = dW_t + \mathrm{mean}(\mu_t(v))dt$ then we have $\mu_t(v) = \mathcal{T}_{v_t}\mathcal{G}_t\mu$.

Now to construct $\mu_t(w)$, let B_t be an independent N-dimensional Brownian motion and construct μ_t and w_t in the same way as stochastic localization driven by Brownian motion B_t, for all times t up to stopping time

$$\tau = \inf\{t \ge 0 : \|v_t\| \ge \|w_t\|\}.$$

Observe by continuity that $\|v_\tau\| = \|w_\tau\|$ almost surely since initially $\|v\| \le \|w\|$. From time τ onward, we define $\mu_t(w)$ in the following way. Let R be an orthogonal operator such that $Rv_\tau = w_\tau$. For $t > \tau$, define

$$B'_t = R(W_t - W_\tau)\mathbb{1}[t > \tau] + B_{\min(t,\tau)}$$

and observe that the law of the process B'_t is that of a standard N-dimensional Brownian motion. Now for all times t we can define

$$d\mu_t(w)(x) = \mu_t(w)(x)\langle x - \mathrm{mean}(\mu_t(w)), dB'_t\rangle$$

to be the stochastic localization process driven by B'_t, which extends the definition we gave before past the stopping time τ. Similarly, we can now define for all times that w_t is the corresponding tilt given by the SDE $dw_t = dB'_t + \mathrm{mean}(\mu_t(w))dt$.

On the other hand, for $t > \tau$ observe that

$$d(Rv_t) = Rdv_t = RdW_t + R\,\mathrm{mean}(\mathcal{T}_{v_t}\mathcal{G}_t\mu) = dB'_t + \mathrm{mean}(\mathcal{T}_{Rv_t}\mathcal{G}_t\mu)$$

by Lemma 63. This is exactly the SDE defining w_t, so by standard uniqueness theory [KS14] we have that $w_t = Rv_t$ for $t > \tau$.

In conclusion, from the definition of the stopping time we have that $\|v_t\| \le \|w_t\|$ for $t < \tau$, and by our construction above we have that $\|v_t\| = \|w_t\|$ for $t \ge \tau$. This proves the final claim. □

Lemma 63. *Suppose that μ is a spherically symmetric distribution in $\mathbb{R}^N$. Then for any orthogonal matrix R and vector $v \in \mathbb{R}^N$ such that $\mathcal{T}_v\mu$ exists, $R\,\mathrm{mean}(\mathcal{T}_v\mu) = \mathrm{mean}(\mathcal{T}_{Rv}\nu)$.*

Proof. Expanding the definition, we have that

$$\begin{aligned} R\,\mathrm{mean}(\mathcal{T}_v\mu) &= \frac{1}{Z_v}\int Rxe^{\langle v,x\rangle}d\mu(x) \\ &= \frac{1}{Z_v}\int ye^{\langle v,R^{-1}y\rangle}d\mu(y) \\ &= \frac{1}{Z_v}\int ye^{\langle Rv,y\rangle}d\mu(y) \\ &= \mathrm{mean}(\mathcal{T}_{Rv}\mu) \end{aligned}$$

where $Z_v = \int e^{\langle v,x\rangle}d\mu(x)$ is the (rotationally invariant) moment generating function, in the second step we made a change of variables $y = Rx$ and used rotational invariance of μ i.e. $d\mu(Rx) = d\mu(x)$, and in the third step we again used that R is an orthogonal transformation i.e. $\langle v, R^{-1}y\rangle = v^\intercal R^{-1}y = v^\intercal R^\intercal y = \langle Rv, y\rangle$. □

5 Rapid mixing of Glauber dynamics in spin systems over semi-log-concave base measures

In this section, we establish a version of our analysis for spin systems where the base measure is semi-log-concave. This greatly generalizes the setting of the Ising models. The general result is generally less tight than specializing the technique to the particular measure of interest, but as we will show the result is easy to apply, and interestingly the resulting integral equation often admits a closed-formula solution.

We will need to use the following standard notation in the argument: $A \otimes B$ denotes the Kronecker product of matrices A and B. Explicitly, if A is a matrix of dimension $m \times n$ then

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1n}B \\ \vdots & \ddots & \vdots \\ A_{m1}B & \cdots & A_{mn}B \end{bmatrix}.$$

Theorem 64. *Suppose that $\mu = \bigotimes_{i=1}^n \mu^{(i)}$ is a product measure where each $\mu^{(i)}$ is a probability measure on $\mathbb{R}^N$. Let $J \succeq 0$ with $J_{ii} = \alpha$ for all $i \in [n]$, and define the measure ν by its density*

$$\frac{d\nu}{d\mu}(x) \propto \exp\left(\frac{1}{2}\sum_{i,j} J_{ij}\langle x_i, x_j\rangle\right).$$

Let $\eta = \alpha/\|J\|_{OP}$. Suppose that for all $i \in [n]$ and $s \in [0, \alpha]$ the measure $\mathcal{G}_{-s}\mu^{(i)}$ defined by

$$\frac{d\mathcal{G}_{-s}\mu^{(i)}}{d\mu^{(i)}}(x) \propto \exp(s\|x\|^2/2) \tag{14}$$

exists and is $\rho(s) > 0$ semi-log-concave. Then $\|\mathrm{cov}(\nu)\|_{OP} \le q_{\eta,\rho}(\|J\|_{OP})$ where $q_{\eta,\rho}$ solves the integral equation

$$q_{\eta,\rho}(z) = \frac{1}{\eta z + [\rho(\eta z)]^{-1}} + \int_0^z q_{\eta,\rho}(s)^2 ds \tag{15}$$

and ATE holds with constant at most

$$\exp\left(\int_0^{\|J\|_{OP}} q_{\eta,\rho}(z)dz\right).$$

Remark 65. The assumption that the reweighted measure (14) is $\rho(s)$-semi-log-concave lets us obtain much better results than if we only assume $\rho(0)$-semi-log-concavity for $\mu^{(i)}$. For example, if $\mu^{(i)}$ is γ-strongly log concave, then this assumption is satisfied with $\rho(s) = \frac{1}{\gamma - s}$ by the Brascamp-Lieb theorem [BL02]. In fact, the same bound holds for all γ-semi-log-concave measures, see Theorem 102. In our main applications, ρ is constant, which is much better than the generic guarantee, and also in this case we will show that the integral equation is analytically solvable.

In the next sections, we prove this theorem and also show how to derived the aforementioned analytical solution.

5.1 Decoupling argument for stochastic localization

Our analysis of the Ising model involved a key reduction from arguing about a joint stochastic localization process to a more tractable *single-site* stochastic localization process. As a first step in our argument, we revisit this idea and formalize it as a general decoupling principle.

Lemma 66. *Suppose that $\mu = \bigotimes_{i=1}^n \mu^{(i)}$ is a product measure where each $\mu^{(i)}$ is a probability measure on $\mathbb{R}^N$. Let $J \succeq 0$ with $J_{ii} = \alpha$ for all i, and define the measure ν_0 by its density*

$$\frac{d\nu_0}{d\mu}(x) \propto \exp\left(\frac{1}{2}\sum_{i,j} J_{ij}\langle x_i, x_j\rangle\right).$$

Let ν_t for $t \geq 0$ be the stochastic localization process with driving matrix $C_t = (J \otimes I_N)^{1/2}$, and for $i \in [n]$ let $\nu_t^{(i)} = \mathcal{L}_{X\sim\nu_t}(X_i)$ denote the (random) marginal law of X_i under ν_t. For every $i \in [n]$ there exists a probability measure λ_i on $\mathbb{R}^N$ such that the marginal law of the random measure $\nu_1^{(i)}$ satisfies

$$\mathcal{L}\left(\nu_1^{(i)}\right) = \mathcal{L}\left(\tilde{\nu}_1^{(i)}\right)$$

where the random measure $\tilde{\nu}_1^{(i)}$ is constructed as follows:

1. *Let $H_i \sim \lambda_i$ and $\tilde{W}_i = (\tilde{W}_{i,t})_t$ is a standard Brownian motion in $\mathbb{R}^N$ independent of H_i.*
2. *Let*
$$\frac{d\tilde{\nu}_0^{(i)}}{d\mu^{(i)}}(x_i^*) \propto \exp(J_{ii}\|x_i^*\|^2 + \langle H_i, x_i^*\rangle).$$
3. *Then, $(\tilde{\nu}_t^{(i)})_{t\geq 0}$ is the stochastic localization process generated by $\tilde{W}$ with driving matrix $C_t = J_{ii}I_N$ starting from $\tilde{\nu}_0^{(i)}$.*

Furthermore, we have the following block-diagonal decomposition of the average covariance:

$$\mathbb{E}[\mathrm{cov}(\nu_1)] = \begin{bmatrix} \mathbb{E}\left[\mathrm{cov}\left(\tilde{\nu}_1^{(1)}\right)\right] & & \\ & \ddots & \\ & & \mathbb{E}\left[\mathrm{cov}\left(\tilde{\nu}_1^{(n)}\right)\right] \end{bmatrix}.$$

Proof. The first step is to apply Theorem 32 — let B_t and W_t be the Brownian motions defined there. By Theorem 32, at time one stochastic localization cancels out the quadratic term in the log-density yielding

$$\frac{d\nu_1}{d\mu}(x) \propto \exp\big(\langle (J \otimes I_N) y_1, x\rangle\big)$$

where $y_1 = X^* + (J \otimes I_N)^{-1/2} B_1 \in \mathbb{R}^{Nn}$ and $X^* \sim \nu_0$. Defining $h^1 = (J \otimes I_N) y_1$ and writing $h^1 = (h_i^1)_{i=1}^n$ we have that for any $i \in [n]$

$$h_i^1 = \sum_{j=1}^n J_{ij} X_j^* + \sqrt{J_{ii}} G_i$$

where $G_i = \frac{(B_1)_i}{\sqrt{J_{ii}}} \sim \mathcal{N}(0, I_N)$ is independent of X^*. Furthermore,

$$\frac{d\nu_1^{(i)}}{\mu^{(i)}}(x_1) \propto \exp(h_i^1 x_1).$$

Define $H_i = \sum_{j \neq i} J_{ij} X_j^*$ so that

$$h_i^1 = H_i + J_{ii} X_i^* + \sqrt{J_{ii}} G_i.$$

Let $\tilde{\nu}_0^{(i)}$ denote the conditional law of X_i^* given H_i and observe that its relative density with respect to the base measure is

$$\frac{d\tilde{\nu}_0^{(i)}}{d\mu^{(i)}}(x_i^*) \propto \exp\left(J_{ii} \|x_i^*\|^2 + \langle H_i, x_i^* \rangle\right).$$

Let $\tilde{\nu}_t^{(i)}$ be defined as the stochastic localization process with driving matrix $\tilde{C}_t = J_{ii} I_N$ generated by a fresh N-dimensional Brownian motion $\tilde{W}_i$. By appealing to Theorem 32 again to determine the law of $\tilde{\nu}_1^{(i)}$ conditional on H_i, we find that the two constructions of random measures on spin i via stochastic localization have the same conditional law:

$$\mathcal{L}(\nu_1^{(i)} \mid H_i) = \mathcal{L}(\tilde{\nu}_1^{(i)} \mid H_i).$$

Taking the expectation over H_i yields the stated equality in law, where λ_i is the law of H_i. From this, the equality in covariances follows since ν_1 is a product measure. □

5.2 Proof of covariance bound and approximate tensorization

The decoupled stochastic localization process can be controlled using the following lemma.

Lemma 67 (Eqn. (16) of [Eld20])**.** *Let μ_t be the stochastic localization process initialized at μ_0 with time-homogeneous driving matrix $C_t = Q$. Then*

$$\mathbb{E}[Q \operatorname{cov}(\mu_t) Q] \preceq \left(tI + (Q \operatorname{cov}(\mu_0) Q)^{-1}\right)^{-1}$$

Note that this result upper bounds the average covariance matrix at time t by a function of its covariance at time 0 — so in a sense, if the trickledown concept is based upon backwards induction, this inequality is a complementary application of forward induction. The lemma has a short proof by combining Gronwall's inequality with the SDE for the covariance matrix. See also [EM22] for a related proof of this lemma without direct reference to stochastic localization.

Proof of Theorem 64. Some steps in this proof are similar to the analysis for the Ising model, so we elaborate in the most detail on the parts which differ. We view the spin system as a distribution on $\mathbb{R}^{Nn}$, and let ν_t be the stochastic localization process with the time-invariant driving matrix $C_t = (J \otimes I_N)^{1/2}$. By Theorem 32, or by a direct computation, this yields

$$\frac{d\nu_t}{d\mu}(x) \propto \exp\left(\langle h_t, x_i\rangle + \frac{1-t}{2}\sum_{i,j} J_{ij}\langle x_i, x_j\rangle\right)$$

for some adapted process h_t in $\mathbb{R}^{Nn}$. Recall the trickle-down equation

$$\operatorname{cov}(\nu_t) = \mathbb{E}[\operatorname{cov}(\nu_1) \mid \mathcal{F}_t] + \int_t^1 \mathbb{E}[\operatorname{cov}(\nu_s) J \operatorname{cov}(\nu_s) \mid \mathcal{F}_t] ds$$

and applying the triangle inequality yields

$$\|\operatorname{cov}(\nu_t)\|_{OP} \le \|\mathbb{E}[\operatorname{cov}(\nu_1) \mid \mathcal{F}_t]\|_{OP} + \int_t^1 \|\mathbb{E}[\operatorname{cov}(\nu_s) J \operatorname{cov}(\nu_s) \mid \mathcal{F}_t]\|_{OP} ds.$$

By the decoupling argument from Lemma 66, the covariance matrix of ν_1 is block-diagonal, and each block can be reinterpreted as the covariance resulting from running stochastic localization with identity covariance in $\mathbb{R}^N$ for time $J_{ii}(1-t)$ starting from the measure $\mathcal{G}_{-J_{ii}(1-t)}\mu^{(i)}$. By Lemma 67 and the assumption that $J_{ii} = \alpha$ we get that for each i, the corresponding block of the covariance satisfies

$$\mathbb{E}[\operatorname{cov}(\nu_1(X_i = \cdot)) \mid \mathcal{F}_t] \le \left[\alpha(1-t)I + \operatorname{cov}(\nu_t(X_i = \cdot))^{-1}\right]^{-1} \le \frac{1}{\alpha(1-t) + 1/\rho(\alpha(1-t))} I$$

where in the last step we used the semi-logconcavity assumption. We therefore have the inequality

$$\|\operatorname{cov}(\nu_t)\|_{OP} \le \frac{1}{\alpha(1-t) + 1/\rho(\alpha(1-t))} + \|J\|_{OP}\int_t^1 \mathbb{E}[\|\operatorname{cov}(\nu_s)\|_{OP}^2 \mid \mathcal{F}_t] ds.$$

As in the argument for the Ising model, by considering at every time the worst-case value of $\|\operatorname{cov}(\nu_t)\|_{OP}$ and making a comparison argument, this yields the inequality $\|\operatorname{cov}(\nu_t)\|_{OP} \le f(t)$ where f solves the integral equation

$$f(t) = \frac{1}{\alpha(1-t) + 1/\rho(\alpha(1-t))} + \|J\|_{OP}\int_t^1 f(s)^2 ds.$$

Recalling that $\eta = \alpha/\|J\|_{OP}$, so $\alpha = \eta\|J\|_{OP}$, we can rewrite this as

$$f(t) = \frac{1}{\eta\|J\|_{OP}(1-t) + 1/\rho(\alpha(1-t))} + \int_0^{\|J\|_{OP}(1-t)} f(t + s/\|J\|_{OP})^2 ds.$$

Letting $z = \|J\|_{OP}(1-t)$ and $q_{\eta,\rho}(z) = f(1 - z/\|J\|_{OP})$ yields

$$\begin{aligned} q_{\eta,\rho}(z) &= f(1 - z/\|J\|_{OP}) \\ &= \frac{1}{\eta z + [\rho(\eta z)]^{-1}} + \int_0^z f(1 - (z-s)/\|J\|_{OP})^2 ds \\ &= \frac{1}{\eta z + [\rho(\eta z)]^{-1}} + \int_0^z q_{\eta,\rho}(z-s)^2 ds \\ &= \frac{1}{\eta z + [\rho(\eta z)]^{-1}} + \int_0^z q_{\eta,\rho}(s)^2 ds. \end{aligned}$$

Approximate Tensorization of Entropy (ATE) follows from the covariance bound in the same way as in the Ising case (using entropic stability and the super-martingale property). □

5.2.1 Aside: A general distribution-specific bound

In some cases we expect to obtain tighter results by avoiding the generic bound from Lemma 67, e.g. as we did in the Ising case. In this spirit and for completeness, we record the following result which follows from the ideas we have already explained. Although computing the bound reduces to evaluating quantities which are independent of the number of spins n, in practice it may not be as easy to apply as the bound based upon semi-log-concavity. We do not build upon this result in the remainder of this paper.

Theorem 68. *Suppose $N \geq 2, n \geq 1$ and that ν is the probability distribution on $\mathbb{R}^{Nn}$ with density*

$$\frac{d\nu}{d\mu}(x) \propto \exp\left(\frac{1}{2}\sum_{1\leq i,j\leq n} J_{ij}\langle x_i, x_j\rangle + \sum_{i=1}^{n}\langle h_i, x_i\rangle\right)$$

with respect to the probability measure $\mu = \otimes_{i=1}^{n}\mu_0$ where μ_0 is a probability measure on $\mathbb{R}^N$. Here J and h are parameters and we assume the interaction matrix J satisfies $J \geq 0$. Suppose, without loss of generality, that the diagonal of J is constant, so there exists α such that $J_{ii} = \alpha \geq 0$. Let $\eta = \alpha/\|J\|_{\mathrm{op}} \in [0,1]$. Then

$$\|\mathrm{cov}(\nu)\|_{\mathrm{op}} \leq q_{\eta,\mu_0}(\|J\|_{\mathrm{op}})$$

where $q_{\eta,\mu_0} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ solves the Volterra integral equation

$$q_{\eta,\mu_0}(z) = r_{\mu_0}(\eta z) + \int_0^z q_{\eta,\mu_0}(y)^2 dy,$$

where

$$r_{\mu_0}(t) = \sup_{b\in\mathbb{R}^N}\left\|\mathbb{E}_{x\sim\mu_0, g\sim\mathcal{N}(0,I)}\left[\mathrm{cov}\left(\mathcal{G}_{-t}\mathcal{T}_{b+tx+\sqrt{t}g}\mu_0\right)\right]\right\|_{\mathrm{op}}.$$

Furthermore, Approximate Tensorization of Entropy (ATE) is satisfied with constant at most

$$\exp\left(\int_0^{\|J\|_{\mathrm{op}}} q_{\eta,N}(z)dz\right).$$

Proof. The proof is the same as that of Theorem 64, except that we do not apply Lemma 67 in which case the quantity r_{μ_0} remains in the final bound. □

5.3 Analytical solution using the Riccati equation

The following argument allows us to explicitly solve the integral equation in many cases of interest, where the semi-logconcavity bound is constant.

Theorem 69. *Suppose that the assumption of Theorem 64 is satisfied for a constant function $\rho(s) = \rho > 0$. Then the solution of (15) exists, and is finite and unique, for $z \in [0, s(\eta)/\rho)$ and is given by*

$$q_{\eta,\rho}(z) = \rho Q_\eta(\rho z)$$

where $Q_\eta : [0, s(\eta)) \to \mathbb{R}_{\geq 0}$ is given by

$$Q_\eta(z) = \eta\frac{-\lambda_1\lambda_2^2(\eta z+1)^{\lambda_1-\lambda_2} + \lambda_1^2\lambda_2}{\lambda_2^2(\eta z+1)^{\lambda_1-\lambda_2+1} - \lambda_1^2(\eta z+1)},$$

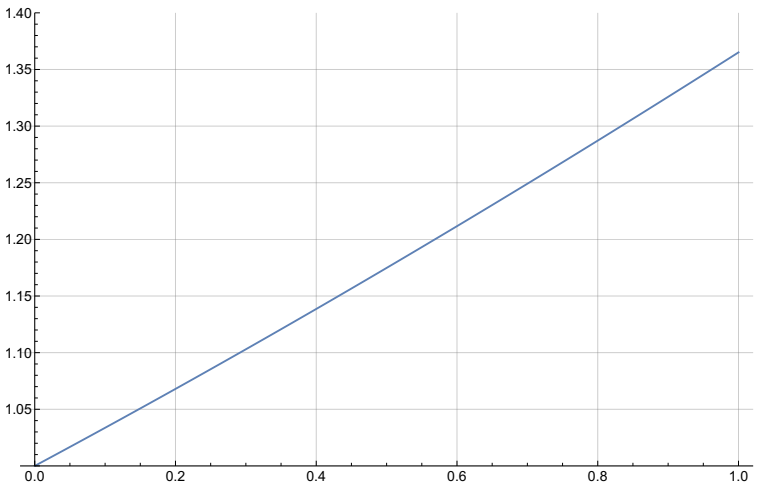

Figure 3: Plot of $s(\eta)$ as a function of η varies from 0 to 1. The function is very close to, but not exactly, affine on this domain.

with $\lambda_1 > 0 > \lambda_2$ the two roots of the quadratic equation $\eta\lambda^2 - \eta\lambda - 1 = 0$, explicitly

$$\lambda_1 = \lambda_1(\eta) = \frac{\eta + \sqrt{\eta^2 + 4\eta}}{2\eta}, \qquad \lambda_2 = \lambda_2(\eta) = \frac{\eta - \sqrt{\eta^2 + 4\eta}}{2\eta},$$

and where

$$s(\eta) = \frac{1}{\eta}\left[\left|\frac{\lambda_1}{\lambda_2}\right|^{2/(\lambda_1 - \lambda_2)} - 1\right].$$

Proof. In this case we have

$$q_{\eta,\rho}(z) = \frac{1}{\eta z + \rho^{-1}} + \int_0^z q_{\eta,\rho}(s)^2 ds. \tag{16}$$

We now show how to exhibit a solution of this equation (we already know it will be unique, see Lemma 55). Differentiating with respect to z yields

$$q'_{\eta,\rho}(z) = \frac{-\eta}{(\eta z + \rho^{-1})^2} + q_{\eta,\rho}(z)^2.$$

This is a Riccati equation so we can solve it by making an appropriate substitution (see e.g. [Rei72]). Explicitly, making the substitution $q_{\eta,\rho} = -u'/u$ yields a differential equation

$$\frac{d^2u}{dz^2} - \frac{\eta u}{(\eta z + \rho^{-1})^2} = 0.$$

Making the substitution $x = \log(\eta\rho z + 1)$ and simplifying yields the differential equation

$$\eta u'' - \eta u' - u = 0.$$

We can guess particular solutions by observing that $u = e^{\lambda x}$ yields the quadratic equation

$$\eta\lambda^2 - \eta\lambda - 1 = 0$$

which has solutions

$$\lambda = \frac{\eta \pm \sqrt{\eta^2 + 4\eta}}{2\eta}$$

and so in general for some $c_1, c_2 \in \mathbb{R}$ the solution must be of the form

$$u = c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x} = c_1(\eta\rho z + 1)^{\lambda_1} + c_2(\eta\rho z + 1)^{\lambda_2}$$

where $\lambda_1 > 0, \lambda_2 < 0$ are the two roots. Substituting back, this gives

$$\begin{aligned} q_{\eta,\rho} &= \eta\rho \, \frac{-c_1\lambda_1(\eta\rho z + 1)^{\lambda_1 - 1} - c_2\lambda_2(\eta\rho z + 1)^{\lambda_2 - 1}}{c_1(\eta\rho z + 1)^{\lambda_1} + c_2(\eta\rho z + 1)^{\lambda_2}} \\ &= \eta\rho \, \frac{-c_1\lambda_1(\eta\rho z + 1)^{\lambda_1 - \lambda_2} - c_2\lambda_2}{c_1(\eta\rho z + 1)^{\lambda_1 - \lambda_2 + 1} + c_2(\eta\rho z + 1)}. \end{aligned}$$

Assume for the purpose of contradiction that $c_1 = 0$, then $q_{\eta,\rho}(z) = -\eta\rho\lambda_2/(\eta\rho z + 1)$ so $q_{\eta,\rho}$ is a solution for all $z \geq 0$ and $q_{\eta,\rho} \to 0$ as $z \to \infty$. However, from (16) we can see $q_{\eta,\rho}(0) = \rho > 0$, so the integral on the right hand side of Eq. (16) and thereby $q_{\eta,\rho}$ are bounded below by a positive constant for all z. By contradiction, $c_1 \neq 0$.

Therefore, if we define $C = -c_2/c_1$ we can write

$$q_{\eta,\rho}(z) = \eta\rho \frac{-\lambda_1(\eta\rho z + 1)^{\lambda_1 - \lambda_2} + C\lambda_2}{(\eta\rho z + 1)^{\lambda_1 - \lambda_2 + 1} - C(\eta\rho z + 1)}. \tag{17}$$

In the case $z = 0$, we therefore have

$$\rho = q_\eta(0) = \eta\rho \frac{-\lambda_1 + C\lambda_2}{1 - C}$$

so $1 - C = -\eta\lambda_1 + C\eta\lambda_2$, i.e.

$$C = \frac{1 + \eta\lambda_1}{1 + \eta\lambda_2} = \frac{\lambda_1^2}{\lambda_2^2}$$

where we used that λ_1, λ_2 are roots of $\eta\lambda^2 = 1 + \eta\lambda$. So multiplying through by λ_2^2, we have

$$q_{\eta,\rho}(z) = \eta\rho \frac{-\lambda_1\lambda_2^2(\eta\rho z + 1)^{\lambda_1 - \lambda_2} + \lambda_1^2\lambda_2}{\lambda_2^2(\eta\rho z + 1)^{\lambda_1 - \lambda_2 + 1} - \lambda_1^2(\eta\rho z + 1)}$$

This formula is only valid for z from 0 to the first positive root z_1 of the denominator (since e.g. the right hand side is typically infinite and therefore no longer a solution of the ODE at time z_1). Since $\eta\rho z + 1 > 0$ for $z \geq 0$, the positive roots must satisfy

$$(\eta\rho z + 1)^{\lambda_1 - \lambda_2} = C.$$

So in fact there is exactly one positive root and it is given by

$$\frac{1}{\eta\rho}\left[C^{1/(\lambda_1 - \lambda_2)} - 1\right] = s(\eta)/\rho.$$

□

6 Rapid mixing of Langevin and Glauber dynamics in $O(N)$ model

For any $N \geq 1$, let S_{N-1} be the unit sphere in $\mathbb{R}^N$. Recall that for $N \geq 2$, this is a connected smooth submanifold of $\mathbb{R}^N$ with dimension $N-1$.

6.1 Covariance bounds and rapid mixing of Langevin

We start with the following result which gives the semi-log-concavity estimate.

Theorem 70 (Theorem D.2 of [DLS78])**.** *For $N \geq 1$, define $f_N(h)$ to be the cumulant generating function for the $N-1$ dimensional unit sphere, so*

$$f_N(h) = \log \int_{s \in S_{N-1}} \exp(h \cdot s) \mu(dA)$$

where μ is the uniform measure on the sphere. Then:

1. *f is spherically symmetric, so $f_N(h) = F_N(\|h\|_2)$ where $F_N : \mathbb{R}_{\geq 0} \to \mathbb{R}$.*
2. *The function $g_N(h) = F_N'(h)$ solves the differential equation*

$$g_N'(h) + g_N(h)^2 + (N-1) g_N(h)/h = 1$$

   on $\mathbb{R}_{>0}$ with boundary conditions $g_N(0) = 0$ and $g_N'(0) = 1/N$.
3. *We have $g_N''(0) = 0$ and $g_N''(h) < 0$ for all $h > 0$. In particular, g_N is concave on $\mathbb{R}_{\geq 0}$.*
4. *For $h \neq 0$, the eigenvalues of $\nabla^2 f_N(h)$ are $g_N'(\|h\|_2)$ with multiplicity one and $g_N(\|h\|_2)/\|h\|_2$ with multiplicity $N-1$. In particular, for all h we have $\|\nabla^2 f_N\|_{\text{op}} \leq \max_{h \geq 0} g_N'(h) = 1/N$.*

Remark 71. The differential equation defining g_N is a Riccati equation which actually admits an explicit solution in terms of Bessel functions. The solution can be found using the same techniques from the proof of Theorem 69.

Theorem 72. *Suppose $N \geq 2, n \geq 1$ and that ν is the probability distribution on $S_{N-1}^{\otimes n}$ with probability density*

$$\frac{d\nu}{d\mu}(x) \propto \exp\left(\frac{1}{2} \sum_{1 \leq i,j \leq n} J_{ij} \langle x_i, x_j \rangle + \sum_{i=1}^n \langle h_i, x_i \rangle \right)$$

with respect to the uniform measure μ on $S_{N-1}^{\otimes n}$. Here J, h are parameters and we assume the interaction matrix J satisfying $J \geq 0$. Suppose (without loss of generality) that the diagonal of J is constant, so there exists α such that $J_{ii} = \alpha \geq 0$. Let $\eta = \alpha / \|J\|_{\text{op}} \in [0,1]$.

Then

$$\|\text{cov}(\nu)\|_{\text{op}} \leq q_{\eta, 1/N}(\|J\|_{\text{op}})$$

where $q_{\eta,1/N}$ is as defined in Theorem 69. Furthermore, the log-Sobolev constant of the Langevin dynamics is at least

$$C_N \exp\left(- \int_0^{\|J\|_{\text{op}}} q_{\eta,1/N}(z) dz \right)$$

where $C_N > 0$ is a constant depending only on N, inherited from [ZQM11].

Proof. The covariance bound follows from Theorem 64, Theorem 69, and the fact that the spherical measure is $1/N$-semi-log-concave from Theorem 70.

The log-Sobolev bound also follows from an completely analogous argument (using the same stochastic localization process as in Theorem 64, entropic stability, and the fact that the right hand side of (5) is a martingale under stochastic localization), except that to cover the base case we appeal to the uniform log-sobolev inequality from [ZQM11]. □

6.2 Nearly linear time sampling via Glauber dynamics

Theorem 73. *In the same setting as Theorem 72, the probability measure ν satisfies ATE with constant at most*

$$\exp\left(\int_0^{\|J\|_{\mathrm{op}}} q_{\eta,1/N}(z)dz\right).$$

Furthermore, a single step of the Glauber dynamics can be ϵ-approximately implemented using $O(\log(1/\epsilon)(\log n + \log\log(1+B))$ arithmetic operations, where $B = \max_i \sum_j |J_{ij}| + \|h_i\|$. In other words, we can sample from a distribution that is ϵ-close in total variation distance to the conditional distribution at the spin chosen to be updated. Therefore, to sample from a distribution ε-close to ν in total variation distance, we run the Glauber dynamics for $T = n\log(n/\varepsilon)$ steps and set $\epsilon = \frac{\varepsilon}{T}$. The total runtime is $O(n\log(n/\varepsilon)(\log N + \log\log(1+B)))$.

Proof. ATE follows from Theorem 69 given the semi-log-concavity bound previously established. What remains is to discuss the implementation of Glauber dynamics.

Each step of the Glauber dynamics sample the spin $x_i \in S^{N-1}$ conditional on the remaining spins x_j being set at u_j. The density $\pi(x_i)$ at x_i is $\exp(\langle h_i + \sum J_{ij}u_j, x_i\rangle)$. By applying a suitable rotation matrix, we can assume without loss of generality that

$$h_i + \sum J_{ij}u_j = be_1$$

where $b \in \mathbb{R}_{\geq 0}$ and e_1 is the basis vector with the first coordinate being 1 and the rest of the coordinates being 0. By the triangle inequality, we have $b \leq B$ where B is as in the theorem statement.

If $b = 0$ then x_i is uniformly distributed on the sphere S^{N-1}, thus we can sample x_i in $O(n)$ time. Below, suppose $b > 0$. If $N = 1$ then π is a Bernoulli distribution, so sampling can be done in $O(1)$ time. (This is the case of the Ising model.) Below, we assume $n \geq 2$.

Let $y = \langle x_i, e_1\rangle$, then $y \in [-1, 1]$. Note that if we can sample y exactly, then to sample x_i, we only need to uniformly sample from the set $\{u \in S^{N-1}\langle u, e_1\rangle = y\}$, which is isomorphic to the sphere S^{n-2}, and this can be done in $O(N)$ time.

By Fact 74, the density of y induced by π is

$$\pi(y) \propto \rho(y) := \exp(by)(1-y^2)^{(N-3)/2}.$$

Case $N = 2$. If $N = 2$, then we can do a change of variable by writing $y = \cos(\theta)$ for $\theta \in [-\pi, \pi)$. Then ρ induces a density φ supported on $[-\pi, \pi)$ defined by $\varphi(\theta) = \exp(b\cos(\theta))$. We consider the restriction $\tilde{\varphi}$ of φ to $[-\pi/4, \pi/4)$, which satisfies the preconditions of Lemma 75 with $V(\theta) = b(1-\cos(\theta))$ and $V''(\theta) = b\cos(\theta) \in [b/\sqrt{2}, b]$ for $\theta \in [-\pi/4, \pi/4)$. Next, we sample from φ by rejection sampling from $\tilde{\varphi}$:

- Sample u uniformly from $\{0,1,2,3\}$
- Sample $\tilde{\theta} \in [-\pi/4, \pi/4)$ from $\tilde{\varphi}$ then output $\theta = \tilde{\theta} + u\pi/2$ with probability $\exp(b(\cos(\theta) - \cos(\tilde{\theta})) \leq 1$.

This rejection sampling algorithm succeeds with probability $\frac{1}{4}\sum_{u=0}^{3}\int_{-\pi/4}^{\pi/4}\exp(b(\cos(\tilde{\theta}+u\pi/2) - \cos(\tilde{\theta}))d\tilde{\theta} \geq 1/4$. Combined this with Lemma 75, we have an algorithm that runs in $O(\log(1/\epsilon))$ time and output a sample from $\hat{\varphi}$ s.t. $d_{TV}(\hat{\varphi}, \varphi) \leq \epsilon$.

Case $N = 3$. If $N = 3$, then the cumulative distribution of ρ can be exactly computed:

$$F(y) = \int_{-1}^{y} \rho(u)du = b^{-1}(\exp(by) - \exp(-b))$$

thus one can sample from ρ in $O(1)$ time by inverse-CDF sampling. Explicitly, we uniformly sample r from $[0, F(1)]$ and then output the unique y s.t. $r = F(y)$.

Remaining cases ($N \geq 4$). We now deal with the four and higher-dimensional cases. Note that $\rho(y) = \exp(-V(y))$ with

$$V(y) = -by - \frac{N-3}{2}\log(1-y^2)$$

and observe that

$$\nabla V(y) = -b + (N-3)\frac{y}{1-y^2}, \qquad \nabla^2 V(y) = \frac{(N-3)(1+y^2)}{(1-y^2)^2} \geq (N-3) > 0.$$

Let $c = \frac{N-3}{b}$. Observe that the unique solution to $\nabla V(y) = 0$ in $[-1,1]$ is $y_0 = \frac{\sqrt{c^2+4}-c}{2} \in (0,1]$. Let

$$\tilde{V}(y) = V(\frac{y}{\sqrt{N-3}} + y_0) - V(y_0)$$

and observe that $\tilde{V}$ is supported on $[\sqrt{N-3}(-1-y_0), \sqrt{N-3}(1-y_0)]$, $\tilde{V}(0) = \tilde{V}'(0) = 0$ and $\tilde{V}''(0) \geq 1$. Note that if $y^2 \geq 1 - \exp(-(\frac{1}{2} + b + V(y_0))) = 1 - (1-y_0^2)^{b(1-y_0)+1/2}$ then

$$V(y) - V(y_0) \geq -\log(1-y^2) - b - V(y_0) \geq \frac{1}{2}$$

and if $y^2 \leq 1 - \exp(-(\frac{1}{2} + b + V(y_0)))$ then $V''(y) \leq 2(N-3)\exp(2(\frac{1}{2} + b + V(y_0)))$.

Thus, Lemma 75 applies for $\tilde{V}$ with $a = \sqrt{N-3}(-1-y_0)$, $b = \sqrt{N-3}(1-y_0)$, and

$$\frac{a'}{\sqrt{N-3}} = -\sqrt{1-(1-y_0^2)^{b(1-y_0)+1/2}} - y_0,$$
$$\frac{b'}{\sqrt{N-3}} = \sqrt{1-(1-y_0^2)^{b(1-y_0)+1/2}} - y_0,$$
$$\kappa = 2\exp(2(\frac{1}{2} + b + V(y_0))).$$

Note that

$$0 \leq 2b(1-y_0) = 2(b - \frac{\sqrt{(N-3)^2+4b^2} - (N-3)}{2}) \leq (N-3)$$

and

$$\frac{1}{1-y_0^2} = \frac{\sqrt{c^2+4}}{2c} + \frac{1}{2} \le \frac{1}{c} + 1 = \frac{b}{N-3} + 1,$$

thus

$$\begin{aligned}\log\log(\max\{|a|,|b|\}\kappa) &\le \log\log(N\exp(1+b+V(y_0))) \\ &= \log\log\left(N\left(\frac{1}{1-y_0^2}\right)^{2b(1-y_0)+1}\right) \\ &= \log\left(\log N + N\log\frac{1}{1-y_0^2}\right) = O(\log N + \log\log(1+b))\end{aligned}$$

where the second line follows from the definition of V. □

Fact 74 (Well-known, see e.g. [ZQM11])**.** *Let $p(x_1)$ be the probability density of the first coordinate X_1 of a random vector X sampled uniformly at random from the $N-1$ dimensional unit sphere S^{N-1} in $\mathbb{R}^N$ centered at the origin. Then p is supported on $[-1,1]$ with density*

$$p(x_1) \propto (1-x_1^2)^{(N-3)/2}.$$

In the implementation of each Glauber dynamics step, we need to sample from a distribution supported on the sphere S^{n-1}, which can be reduced to sampling from a log-concave distribution supported on a real interval. The following shows how to do the latter task using a modification of the technique in [Che+21]. While [Che+21] requires the potential function(s) to be supported on the entire real line and to have uniformly bounded second derivative, our result only requires bounded second derivative in a large enough interval of the support.

Lemma 75. *Consider $V : [a,b] \to \mathbb{R} \cup \{\pm\infty\}$ with $0 \in [a,b]$. Suppose $V(0) = \nabla V(0) = 0$ and $V''(x) \ge \alpha > 0$ for $x \in (a,b)$. There exists a', b' s.t. $a \le a' \le 0 \le b' \le b$ s.t. $V(x) \ge 1/2$ if $x \in [a,a'] \cup [b',b]$ and $V''(x) \le \beta$ if $x \in [a',b']$. Let π be a distribution $[a,b]$ s.t. $p(x) \propto \exp(-V(x))$. For $\kappa = \beta/\alpha$, we have that:*

- *There exists an algorithm which runs in $O(\log\log(\max\{|a|,|b|\}\kappa))$ time and with probability $\Omega(1)$, outputs a sample from p.*
- *Given any $\epsilon > 0$, there exists an algorithm which samples from $\hat{p}$ s.t. $d_{TV}(p,\hat{p}) \le \epsilon$ in $O(\log(\frac{1}{\epsilon}) \cdot \log\log(\max\{|a|,|b|\}\kappa))$ time.*

Proof. Without loss of generality, we can assume $\alpha = 1$ so that $V''(x) \ge 1 \forall x \in [a,b]$ and $V''(x) \le \kappa$ for $x \in [a',b']$. Otherwise, we can replace V with $\tilde{V}$ where $\tilde{V}(x) = V(x/\sqrt{\alpha})$.

We use a rejection sampling algorithm similar to in [Che+21]. To account for the fact that $V''(x)$ is only bounded above by κ in the interval $[a',b']$, we only need to slightly modify the proof so that $x_-, x_+ \in [a',b']$.

First, we build the envelope q, which is a distribution over $[a,b]$, then we draw a sample from q and accept with probability $\frac{p(x)}{q(x)}$. We only need to show that the acceptance probability of this algorithm is $\Omega(1)$. For the second claim, we repeat this algorithm $O(\log(1/\epsilon))$ times and output the result of the first successful run. If all runs fail then output a uniform sample from $[a,b]$. We define q as follows.

1. Find the first index $i_+ \in \{0, 1, \ldots, \lfloor \log_2(b'\sqrt{\kappa}) \rfloor\}$ s.t. $V(2^{i_+}/\sqrt{\kappa}) \geq 1/2$. If i_+ exists, let $x_+ = 2^i_+/\sqrt{\kappa}$ else let $x_+ = b'$
2. Find the first index $i_- \in \{0, 1, \ldots, \lceil \log_2(-a'\sqrt{\kappa}) \rceil\}$ s.t. $V(-2^{i_-}/\sqrt{\kappa}) \geq 1/2$. If i_- exists, let $x_- = -2^i_-/\sqrt{\kappa}$ else let $x_- = a'$
3. Let $q(x) \propto \tilde{q}(x)$ where

$$\tilde{q}(x) = \begin{cases} \exp(-\frac{x-x_-}{2x_-} - \frac{(x-x_-)^2}{2}) & \text{if } x \in [a, x_-] \\ 1 & \text{if } x \in [x_-, x_+] \\ \exp(-\frac{x-x_+}{2x_+} - \frac{(x-x_+)^2}{2}) & \text{if } x \in [x_+, b] \end{cases}$$

First, observe that $V(x) \geq x^2/2$ for all $x \in \mathbb{R}$ and $V(x) \leq \kappa x^2/2$ for all $x \in [a', b']$ thus $1/2 \leq V(b') \leq \kappa(b')^2/2$ and $1/2 \leq V(a') \leq \kappa(a')^2/2$. Thus $\min\{\log(b'\sqrt{\kappa}), \log(-a'\sqrt{\kappa})\} \geq 0$ and the first two steps are well-defined. Sampling from q can be done in $O(1)$ time, since cumulative distribution of q at x is $\mathbb{P}_{Z\sim\mathcal{N}(0,1)}[u < Z < u']$ where u, u' can be computed in $O(1)$ time given x_-, x_+, a, b, x.

Let $\tilde{p}(x) = \exp(-V(x))$. We only need to check that $\tilde{p}(x) \leq \tilde{q}(x)$ and that there exists absolute constant $C > 0$ s.t.

$$CZ_{\tilde{p}} = \int_a^b \exp(-V(x))dx \geq Z_{\tilde{q}} = \int_a^b \tilde{q}(x)dx.$$

To prove this inequality, by using the argument in [Che+21], we only need to check that $x_+ = O\left(\int_0^{x_+} \tilde{p}(x)dx\right)$ and $-x_- = O\left(\int_{x_-}^0 \tilde{p}(x)dx\right)$. We will show the first inequality; the second is entirely analogous. If i_+ doesn't exist, then since V is increasing in $[0, b]$,

$$V(x_+/2) = V(b'/2) \leq V(2^{\lfloor \log_2(b'\sqrt{\kappa}) \rfloor}/\kappa) < 1/2.$$

If i_+ exists and $i_+ > 0$ then by definition of i_+, $V(x_+/2) < 1/2$. In both cases,

$$\int_0^{x_+} \tilde{p}(x)dx \geq \int_0^{x_+/2} \exp(-1/2)dx = \Omega(x_+).$$

The remaining case is $i_+ = 0$. Note that $x_+ = 1/\sqrt{\kappa} \in (0, b']$ thus

$$\int_0^{x_+} \tilde{p}(x)dx \geq \int_0^{1/\kappa} \exp(-\kappa x^2/2)dx \geq \frac{1}{3\sqrt{\kappa}} = \frac{x_+}{3}.$$

The rest of the proof follows from [Che+21, Proof of Theorem 3], since the upper bound $V''(x) \leq \kappa$ is only used for proving the above claims. The rest of the proof in [Che+21] only uses $V''(x) \geq 1$, which does hold for the entire domain of V.

□

7 Sampling Ising models with the polarized walk

In this section, we establish guarantees for sampling antiferromagnetic Ising models on spectral expanders and related results. Major differences from the previous sections is the use of the tools from geometry polynomials and the closely related *polarized walk* rather than the Glauber dynamics.

7.1 Estimates via geometry of polynomials

Lemma 76. *Suppose that ν is a measure on the hypercube $\{\pm 1\}^n$ of the form*

$$\nu(x) \propto \exp\left(-\frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right)$$

for any $\gamma \geq 0$. The homogenized generating polynomial of ν

$$f(z_1, \ldots, z_n, \bar{z}_1, \ldots, \bar{z}_n) = \sum_x \nu(x) \prod_{i:x_i=1} z_i \prod_{i:x_i=-1} \bar{z}_i$$

is $1/2$-fractionally log concave.

Remark 77. Recall that the corresponding homogenized distribution ν^{hom} is a version of ν supported on $\binom{[n]\cup[\bar{n}]}{n}$. We will not need this fact, but it is worthwhile to note that Lemma 76 implies the 2-step down-up walk on ν^{hom} mixes in $\widetilde{O}(n^2)$ steps and entropy contraction of the down operator $D_{n\to(n-2)}$ w.r.t. ν^{hom}, i.e., for all $\pi : \binom{[n]\cup[\bar{n}]}{n} \to \mathbb{R}_{\geq 0}$

$$\mathcal{D}_{\text{KL}}(\pi D_{n\to(n-2)} \parallel \nu D_{n\to(n-2)}) \leq \left(1 - \frac{2}{n(n-1)}\right) \mathcal{D}_{\text{KL}}(\pi \parallel \nu).$$

$D_{n\to(n-2)}$ is the Markov kernel that maps $S \in \binom{[n]\cup[\bar{n}]}{n}$ to a uniformly random subset of S of size $n-2$. The 2-step down-up walk corresponds to the 2-block Glauber dynamics for ν, which operates by choosing uniformly random $\{i, j\} \in \binom{[n]}{2}$, then resampling the spin at i, j conditioned on the remaining assignments.

Remark 78. This result is sharp. Consider $n = 2, h = \vec{0}$. As $\gamma \to \infty$, ν converges to the uniform distribution over $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. The homogenized generating polynomial converges to $z_1\bar{z}_2 + z_2\bar{z}_1$, which is 1/2-fractionally log-concave; that it is exactly 1/2-spectrally independent can be checked by explicitly computing its influence matrix.

Remark 79. It is interesting to compare Lemma 76 to the NP Hardness result from Appendix H of [KLR22]. There they show that, replacing $(\sum_i x_i)^2 = \langle \vec{1}, x\rangle^2$ with $\langle a, x\rangle^2$ in the definition of $\nu(x)$, for an arbitrary vector $a \in \mathbb{Z}^n$, results in a hard problem. More precisely, for a coming from the number partitioning problem, no polynomial time algorithm can approximately sample from such a model within a TV distance of $1/2$ unless RP = NP. Thus, Lemma 76 cannot be recovered from a generic fact about negatively-spiked rank-one Ising models — its validity has to do with the arithmetic structure of the interactions.

One key to the proof of the result will be some fundamental facts relating the spectral independence and fractional log-concavity of different variants of a probability measure μ, which we establish now.

Lemma 80. *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a probability distribution and $\mu^{comp} : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be the complement of μ, i.e., $\mu^{com}([n] \setminus S) = \mu(S)$. Let $\mu^{\text{hom}} : \binom{[n]\cup[\bar{n}]}{n} \to \mathbb{R}_{\geq 0}$ be the homogenization of μ. Then:*

1. *If μ and μ^{comp} are $\frac{1}{\alpha}$-spectrally independent then μ^{hom} is $\frac{2}{\alpha}$-spectrally independent.*
2. *If μ and μ^{comp} are α-FLC then μ^{hom} is $\alpha/2$-FLC.*

Proof. For a probability distribution $\rho : 2^X \to \mathbb{R}_{\geq 0}$, let $\mathrm{mean}(\rho) \in \mathbb{R}^X_{\geq 0}$ be the vector with $\mathrm{mean}(\rho)_i = \mathbb{P}_\rho[i]$.

We consider $\mu, \mu^{\mathrm{com}}, \mu^{\mathrm{hom}}$ as distributions over $2^{[n]}, 2^{[\bar{n}]}, 2^{[n]\cup[\bar{n}]}$ respectively. Thus $\mathrm{cov}(\mu)$ is indexed by $[n]$, $\mathrm{cov}(\mu^{\mathrm{com}})$ is indexed by $[\bar{n}]$ and $\mathrm{cov}(\mu^{\mathrm{hom}})$ is indexed by $[n] \cup [\bar{n}]$.

Because being FLC is equivalent to being spectrally independent under arbitrary tilts, the second claim follows from the first. Indeed, for $\lambda \in \mathbb{R}^{[n]\cup[\bar{n}]}$, $\lambda * \mu^{\mathrm{hom}} = (\tilde{\lambda} * \mu)^{\mathrm{hom}}$ with $\tilde{\lambda}_i = \lambda_i / \lambda_{\bar{i}}$ and $(\tilde{\lambda} * \mu)^{\mathrm{comp}} = \lambda' * \mu^{\mathrm{comp}}$ with $\lambda'_{\bar{i}} = \frac{1}{\tilde{\lambda}_i}$. By the assumption on μ and μ^{comp}, $\tilde{\lambda} * \mu$ and $(\tilde{\lambda} * \mu)$ are $\frac{1}{\alpha}$-spectrally independent thus the first claim implies $\lambda * \mu^{\mathrm{hom}} = (\tilde{\lambda} * \mu)^{\mathrm{hom}}$ is $\frac{2}{\alpha}$-spectrally independent. Since this is true for all λ, μ^{hom} is $\alpha/2$-fractionally log-concave.

We now proceed to prove the claim about spectral independence. First, note that $\mathrm{cov}(\mu) = \mathrm{cov}(\mu^{\mathrm{comp}})$ and $\mathrm{mean}(\mu^{\mathrm{hom}}) = (\mathrm{mean}(\mu), \mathrm{mean}(\mu^{\mathrm{comp}}))$. Furthermore, we claim that

$$\mathrm{cov}(\mu^{\mathrm{hom}}) = \begin{bmatrix} A & -A \\ -A & A \end{bmatrix}$$

with $A = \mathrm{cov}(\mu)$. To see this, observe that $\mathbb{P}_{\mu^{\mathrm{hom}}}[i] = \mathbb{P}_\mu[i]$, $\mathbb{P}_{\mu^{\mathrm{hom}}}[\bar{i}] = \mathbb{P}_{\mu^{\mathrm{comp}}}[\bar{i}] = 1 - \mathbb{P}_\mu[i]$, and

$$\begin{aligned}
(\mathrm{cov}(\mu^{\mathrm{hom}}))_{\bar{i},\bar{j}} &= \mathbb{P}_{\mu^{\mathrm{hom}}}[\bar{i},\bar{j}] - \mathbb{P}_{\mu^{\mathrm{hom}}}[\bar{i}]\,\mathbb{P}_{\mu^{\mathrm{hom}}}[\bar{j}] \\
&= \mathbb{P}_{S\sim\mu}[i, j \notin S] - \mathbb{P}_{S\sim\mu}[i \notin S]\,\mathbb{P}_{S\sim\mu}[j \notin S] \\
&= \mathbb{P}_{S\sim\mu}[i \notin S] - \mathbb{P}_{S\sim\mu}[i \notin S, j \in S] - \mathbb{P}_{S\sim\mu}[i \notin S](1 - \mathbb{P}_{S\sim\mu}[j \in S]) \\
&= -(\mathbb{P}_{S\sim\mu}[i \notin S, j \in S] - \mathbb{P}_{S\sim\mu}[i \notin S]\,\mathbb{P}_{S\sim\mu}[j \in S] \\
&= -(\mathrm{cov}(\mu^{\mathrm{hom}}))_{\bar{i},j},
\end{aligned}$$

and furthermore

$$\begin{aligned}
-(\mathrm{cov}(\mu^{\mathrm{hom}}))_{\bar{i},j} &= -(\mathbb{P}_{S\sim\mu}[j \in S] - \mathbb{P}_{S\sim\mu}[i \in S, j \in S]) + (1 - \mathbb{P}_{S\sim\mu}[i \in S])\,\mathbb{P}_{S\sim\mu}[j \in S] \\
&= \mathbb{P}_{S\sim\mu}[i \in S, j \in S] - \mathbb{P}_{S\sim\mu}[i \in S])\,\mathbb{P}_{S\sim\mu}[j \in S] \\
&= (\mathrm{cov}(\mu^{\mathrm{hom}}))_{i,j}.
\end{aligned}$$

Next, for any vector $\vec{x} \in \mathbb{R}^{[n]\cup[\bar{n}]}$, we can write $\vec{x} = \vec{x}_0 || \vec{x}_1$ with $\vec{x}_0 \in \mathbb{R}^{[n]}$ and $\vec{x}_1 \in \mathbb{R}^{[\bar{n}]}$. Then

$$\begin{aligned}
&\vec{x}^\intercal \left(\frac{2}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu^{\mathrm{hom}})) - \mathrm{cov}(\mu^{\mathrm{hom}})\right)\vec{x} \\
&\quad= \frac{2}{\alpha}(\vec{x}_0^\intercal \mathrm{diag}(\mathrm{mean}(\mu))\vec{x}_0 + \vec{x}_1^\intercal \mathrm{diag}(\mathrm{mean}(\mu^{\mathrm{comp}}))\vec{x}_1) - (\vec{x}_0 - \vec{x}_1)^\intercal A(\vec{x}_0 - \vec{x}_1) \\
&\quad= 2\vec{x}_0^\intercal \left(\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu) - A\right)\vec{x}_0 + \vec{x}_1^\intercal \left(\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu^{\mathrm{comp}})) - A\right)\vec{x}_1 \\
&\qquad+ 2(\vec{x}_0^\intercal A\vec{x}_0 + \vec{x}_1^\intercal A\vec{x}_1) - (\vec{x}_0 - \vec{x}_1)^\intercal A(\vec{x}_0 - \vec{x}_1) \\
&\quad= 2\vec{x}_0^\intercal \left(\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu)) - A\right)\vec{x}_0 + \vec{x}_1^\intercal \left(\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu^{\mathrm{comp}})) - A\right)\vec{x}_1 + (\vec{x}_0 + \vec{x}_1)^\intercal A(\vec{x}_0 + \vec{x}_1) \geq 0
\end{aligned}$$

where the last inequality follows from the following inequalities:

$$\begin{aligned}
\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu)) &\succeq \mathrm{cov}(\mu) = A, \\
\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu^{\mathrm{comp}})) &\succeq \mathrm{Cov}(\mu^{\mathrm{comp}}) = A, \\
A = \mathrm{cov}(\mu) = \mathbb{E}_\mu[(x - \mathrm{mean}(\mu))(x - \mathrm{mean}(\mu))^\intercal] &\succeq 0.
\end{aligned}$$

Since the above inequality is true for all $\vec{x}$, we conclude that $\frac{2}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu^{\mathrm{hom}})) - \mathrm{cov}(\mu^{\mathrm{hom}}) \succeq 0$ and μ^{hom} is $\frac{\alpha}{2}$-FLC. □

Lemma 81. *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a probability distribution. Then:*

1. *If $\Pi(\mu)$ is $\frac{1}{\alpha}$-spectrally independent then so is μ.*
2. *If $\Pi(\mu)$ is α-fractionally log-concave then so is μ.*

Proof. To prove the second claim, we need to show $\lambda * \mu$ is $\frac{1}{\alpha}$-spectrally independent for all $\lambda \in \mathbb{R}^n_{\geq 0}$. Note that $\Pi(\lambda * \mu) = (\lambda || \mathbf{1}) * \Pi(\mu)$ is $\frac{1}{\alpha}$-spectrally independent as a scaling of $\Pi(\mu)$, so the second conclusion follows from the first one. We now prove the first claim about spectral independence.

We examine the covariance matrix of $\Pi(\mu)$. We show that for $i, j \in X$,

$$(\mathrm{cov}(\Pi(\mu)))_{i,j} = (\mathrm{cov}(\mu))_{i,j} \quad \text{and} \quad \mathrm{mean}(\Pi(\mu))_i = \mathrm{mean}(\mu)_i$$

Indeed,

$$\mathrm{mean}(\Pi(\mu))_i = \sum_{S \subseteq X : i \in S, T \cup Y : |T| = n - |S|} \mu(S \sqcup T) = \sum_{S \subseteq X : i \in S, T \cup Y : |T| = n - |S|} \mu(S) \frac{1}{\binom{n}{|S|}} = \sum_{S \subseteq X : i \in S} \mu(S) = \mathrm{mean}(\mu)_i$$

Similarly,

$$\mathbb{P}_{\Pi(\mu)}[i, j] = \sum_{S \subseteq X : i, j \in S, T \cup Y : |T| = n - |S|} \mu(S \sqcup T) = \sum_{S \subseteq X : i, j \in S} \mu(S) = \mathbb{P}_{\mu}[i, j]$$

and

$$(\mathrm{cov}(\Pi(\mu)))_{i,j} = \mathbb{P}_{\Pi(\mu)}[i, j] - \mathbb{P}_{\Pi(\mu)}[i]\,\mathbb{P}_{\Pi(\mu)}[j] = \mathbb{P}_{\mu}[i, j] - \mathbb{P}_{\mu}[i]\,\mathbb{P}_{\mu}[j] = (\mathrm{cov}(\mu))_{i,j}.$$

$1/\alpha$-spectral independence of $\Pi(\mu)$ means that $0 \preceq (\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\Pi(\mu))) - \mathrm{cov}(\Pi(\mu)))$ thus

$$0 \preceq \left(\frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\Pi(\mu))) - \mathrm{cov}(\Pi(\mu))\right)_{X,X} = \frac{1}{\alpha}\mathrm{diag}(\mathrm{mean}(\mu)) - \mathrm{cov}(\mu).$$

□

Proof of Lemma 76. Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be defined by $\mu(S) = \nu(\chi_S)$ where

$$(\chi_S)_i = \begin{cases} 1 & \forall i \in S \\ -1 & \text{else} \end{cases}.$$

Note that $\nu = \mu^{\mathrm{hom}}$. Let $\alpha = 1$. By Lemmas 81 and 82, μ is α-fractionally log concave. Since the generating polynomial of μ^{comp} is again Π_g, μ^{comp} is α-fractionally log concave. Thus by Lemma 80, $\nu = \mu^{\mathrm{hom}}$ is $\frac{\alpha}{2}$-fractionally log-concave. □

Lemma 82. *For any $n \geq 1$ and $c \geq 0$ the bivariate polynomial*

$$g(x, y) = \sum_{k=0}^{n} \binom{n}{k} e^{-c(k-n/2)^2} x^k y^{n-k}$$

is strongly log-concave. Furthermore, its polarization

$$\Pi_n g(x_1, \dots, x_n, y_1, \dots, y_n) = \sum_{k=0}^{n} \sum_{S \in \binom{[n]}{k}, T \in \binom{[n]}{n-k}} e^{-c(k-n/2)^2/2} x_S y_T \frac{1}{\binom{n}{n-k}},$$

which is a multilinear polynomial, is strongly log-concave.

Proof. The first conclusion is is true because the sequence $(e^{-c(k-n/2)^2})_{k=0}^n$ is log-concave, i.e.

$$e^{-c(k-n/2)^2} \geq \sqrt{e^{-c(k-1-n/2)^2} e^{-c(k+1-n/2)^2}}$$

which follows from the the more general fact that the function e^{-cx^2} on the real-line is log-concave. Given this, strong log-concavity of the bivariate polynomial follows from Proposition 41. Strong log-concavity of the multilinear polynomial follows from polarization (Proposition 43). □

7.2 Trickle-down

Lemma 83. *Suppose that ν is a measure on the hypercube $\{\pm 1\}^n$ of the form*

$$\nu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle - \frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right)$$

for some $J \succeq 0, \gamma \geq 0$ and $h \in \mathbb{R}^n$. If $\|J\|_{\mathrm{op}} < 1/2$ then

$$\|\mathrm{cov}(\nu)\|_{\mathrm{op}} \leq \frac{2}{1 - 2\|J\|_{\mathrm{op}}}.$$

Moreover, if we define ν_t to be the stochastic localization process with driving matrix $C_t = J^{1/2}$, then

$$\|\mathrm{cov}(\nu_t)\|_{\mathrm{op}} \leq \frac{2}{1 - 2(1-t)\|J\|_{\mathrm{op}}}$$

Proof. By Theorem 32 (or by explicit calculation), we have that

$$\nu_1(x) \propto \exp\left(\langle h + H_1, x\rangle - \frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right)$$

for some random vector H_1. By Lemma 76 we know that the measure ν_1 is 1/2-fractionally log-concave, and by Proposition 51 this tells us that the covariance matrix of ν_1 has operator norm at most 2. Hence by (10) and submultiplicativity of the operator norm we have that

$$\|\mathrm{cov}(\nu_t)\|_{\mathrm{op}} \leq 2I + \int_t^1 \mathbb{E}[\|\mathrm{cov}(\nu_s)\|^2 \|J\|_{\mathrm{op}}].$$

Observe that the solution to the differential equation $dy/dt = \|J\|_{\mathrm{op}} y^2$ with initial condition $y(0) = 2$ is $y(t) = \frac{2}{1-2\|J\|_{\mathrm{op}} t}$. So by performing a comparison argument in the same way as in the proof of Theorem 54, we obtain the result.

The second statement follows from the first one and the definition of $\nu_t(x)$, since (by e.g. Theorem 32)

$$\nu_t(x) \propto \exp\left(\frac{1}{2}\langle x, (1-t)Jx\rangle + \langle h + H_1, x\rangle - \frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right)$$

and $\|(1-t)J\|_{\mathrm{op}} = (1-t)\|J\|_{\mathrm{op}}$; □

7.3 Dynamics

Definition 84 (polarized operators)**.** Suppose ν is a probability measure on the hypercube $\{\pm 1\}^n$. Let $\tilde{\nu} : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be defined by $\tilde{\nu}(S) = \nu(\chi_S)$ where $\chi_S \in \{\pm 1\}^n$ is defined so that for each $i \in [n]$,

$$(\chi_S)_i = \begin{cases} 1 & \text{if } i \in S, \\ -1 & \text{otherwise.} \end{cases} \tag{18}$$

Consider the polarization $\Pi(\tilde{\nu})$ of $\tilde{\nu}$. For $x \in \{\pm 1\}^n$ let $x_+ = \{i : x_i = 1\}$. The down operator $D_{n\to(n-1)}$ on $\Pi(\nu)$ corresponds to the following Markov kernel D^{pol} mapping[6] $x \in \{\pm 1\}^n$ to $T \in 2^{[n]}$ with the following transition probabilities:

$$D^{\mathrm{pol}}(x,T) = \begin{cases} \frac{1}{n} & \text{if } T = x_+ \setminus \{i\} \text{ for some } i \in x_+, \\ \frac{n-|x_+|}{n} & \text{if } T = x_+, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

The polarized up-operator, define with respect to ν, maps $T \in 2^{[n]}$ to $x \in \{\pm 1\}^n$ with transition probabilities as follows:

$$U^{\mathrm{pol}}(T,x) = \begin{cases} \frac{1}{n} \times \frac{\nu(x)}{\nu D^{\mathrm{pol}}(T)} & \text{if } T = x_+ \setminus \{i\} \text{ for some } i \in x_+, \\ \frac{n-|T|}{n} \times \frac{\nu(x)}{\nu D^{\mathrm{pol}}(T)} & \text{if } T = x_+, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

The polarized walk on ν is the Markov operator $D^{\mathrm{pol}}U^{\mathrm{pol}}$. Explicitly, its transitions are as follows for $x, x' \in \{\pm 1\}^n$:

$$(D^{\mathrm{pol}}U^{\mathrm{pol}})(x,x') = \begin{cases} \frac{n-|x_+|}{n^2} \times \frac{\nu(x')}{\nu D^{\mathrm{pol}}(x_+)} & \text{if } x'_+ = x_+ \cup \{i\}, i \notin x_+, \\ \frac{n-|x'_+|}{n^2} \times \frac{\nu(x')}{\nu D^{\mathrm{pol}}(x'_+)} & \text{if } x'_+ = x_+ \setminus \{i\}, i \in x_+, \\ \frac{1}{n^2} \times \frac{\nu(x')}{\nu D^{\mathrm{pol}}(x_+ \cap x'_+)} & \text{if } x'_+ = x_+ \setminus \{i\} \cup \{j\}, i \in x_+, j \notin x_+, \\ \left(\frac{n-|x_+|}{n}\right)^2 \times \frac{\nu(x)}{\nu D^{\mathrm{pol}}(T)} & \text{if } x' = x, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Proposition 85. *Suppose that ν is a measure on the hypercube $\{\pm 1\}^n$ of the form*

$$\nu(x) \propto \exp\left(-\frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right).$$

The polarized down operator D^{pol} has $(1-1/n)$-entropy contraction with respect to ν.

[6] We can equivalently view it as a Markov operator on $\{\pm 1\}^n$ if we identify a vector $x \in \{\pm 1\}^n$ and the set x_+.

Proof. Recall that by Proposition 41 the polarization $\Pi(\nu)$ is log-concave, thus $D_{n\to(n-1)}$ has $(1-\frac{1}{n})$-entropy contraction wrt $\Pi(\tilde{\nu})$. For any probability distribution ν' on the hypercube $\{\pm 1\}^n$, we therefore have that

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\nu' D^{\mathrm{pol}} \parallel \nu D^{\mathrm{pol}}) &= \mathcal{D}_{\mathrm{KL}}(\Pi(\nu') D_{n\to(n-1)} \parallel \Pi(\nu) D_{n\to(n-1)}) \\ &\le \left(1-\frac{1}{n}\right) \mathcal{D}_{\mathrm{KL}}(\Pi(\nu') \parallel \Pi(\nu)) \\ &= \left(1-\frac{1}{n}\right) \mathcal{D}_{\mathrm{KL}}(\nu' \parallel \nu).\end{aligned}$$

□

Theorem 86. *Suppose that ν is a measure on the hypercube $\{\pm 1\}^n$ of the form*

$$\nu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle - \frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right)$$

for some $J \succeq 0$, $\gamma \ge 0$, and $h \in \mathbb{R}^n$, and suppose that $\|J\|_{\mathrm{op}} < 1/2$. Then:

1. *The polarized down-operator D^{pol} has $\left(1-\frac{1-2\|J\|_{\mathrm{op}}}{n}\right)$-entropy contraction with respect to ν.*
2. *The polarized walk on ν mixes within ϵ-TV distance in $O(\frac{1}{1-2\|J\|_{\mathrm{op}}} n \log(n/\epsilon))$ steps.*
3. *Let*

$$Q = J - \frac{\gamma}{n}\vec{1}\vec{1}^T$$

and let d be the maximum number of non-diagonal nonzero entries in a row of Q. Each step of the polarized walk can be implemented in $O(d \log n)$ time, so the polarized walk outputs a sample within ϵ-TV distance of ν in total runtime $O\left(\frac{1}{1-2\|J\|_{\mathrm{op}}} nd \log(n) \log(n/\epsilon)\right)$

Proof. First we prove the entropy contraction claim. As in Lemma 83, define ν_t to be the stochastic localization process with driving matrix $C_t = J^{1/2}$, then

$$\nu_1(x) := \nu_1^{(H_1)}(x) \propto \exp\left(\langle h + H_1, x\rangle - \frac{\gamma}{2n}\left(\sum_i x_i\right)^2\right)$$

for some random vector H_1. Let π be the distribution of H_1, we can write

$$\nu = \int \nu_1^{(H_1)} d\pi(H_1).$$

Fix a test function $\varphi : \{\pm 1\}^n \to \mathbb{R}_{>0}$. By Lemma 83 and Lemma 36 ν_t is α_t-entropically stable with respect to $\psi(x,y) = \frac{1}{2}\|J^{1/2}(x-y)\|^2 = \frac{1}{2}\|C_t(x-y)\|^2$ with $\alpha_t = \frac{2\|J\|_{\mathrm{op}}}{1-2(1-t)\|J\|_{\mathrm{op}}}$ thus by Proposition 35,

$$\exp\left(-\int_0^1 \alpha_t dt\right) \mathrm{Ent}_\nu[\varphi] \le \mathbb{E}[\mathrm{Ent}_{x\sim\nu_1}[\varphi(x)]] = \mathbb{E}_{H_1\sim\pi}[\mathrm{Ent}_{x\sim\nu_1^{(H_1)}}[\varphi(x)]]$$

By Proposition 85, D^{pol} has $(1-\kappa)$-entropy contraction wrt ν_1 with $\kappa = 1/n$. Using the equivalent definition of entropy contraction in Definition 25 with $F = D^{\mathrm{pol}}$ and the supermartingale property (Lemma 39), we have

$$\kappa\, \mathbb{E}_{H_1\sim\pi}[\mathrm{Ent}_{x\sim\nu_1^{(H_1)}}[\varphi(x)]] \le \mathbb{E}_{H_1\sim\pi}[\mathbb{E}_{z\sim\nu_1^{(H_1)}} D^{\mathrm{pol}}[\mathrm{Ent}_{X\sim\nu_1^{(H_1)}(\cdot\triangleright z)}[\varphi(X)]]] \le \mathbb{E}_{z\sim\mu D^{\mathrm{pol}}}[\mathrm{Ent}_{X\sim\mu(\cdot\triangleright z)}[\varphi(X)]]$$

thus

$$\frac{1}{n}(1-2\|J\|_{\mathrm{op}})\,\mathrm{Ent}_\nu[\varphi] = \kappa \exp\left(-\int_0^1 \frac{2}{1/\|J\|_{\mathrm{op}} - 2(1-t)}dt\right)\mathrm{Ent}_\nu[\varphi] \le \mathbb{E}_{z\sim\mu D^{\mathrm{pol}}}[\mathrm{Ent}_{X\sim\mu(\cdot\triangleright z)}[\varphi(X)]].$$

So we have shown that D^{pol} has $(1-\frac{1-2\|J\|_{\mathrm{op}}}{n})$-entropy contraction with respect to μ.

Approximate tensorization of entropy implies a mixing time bound which depends on the probability of the smallest atom in the probability measure. We now eliminate this dependence to state a slightly sharper bound.

Exchange inequality. We show $\Pi(\nu)$ satisfy the exchange inequality as defined in [Ana+20, Lemma 23], which implies that the down-up walk on $\Pi(\nu)$ reaches a poly(n)-warm start after $O(n\log n)$ steps and thus the same holds for the polarized walk on ν.

The exchange inequality says that for any sets R, R' in the support of $\Pi(\nu)$ and $u \in R\setminus R'$, there exists $v \in R'\setminus R$ such that

$$\Pi(\nu)(R)\,\Pi(\nu)(R') \le 2^{O(n)}\,\Pi(\nu)(R\setminus\{u\}\cup\{v\})\,\Pi(\nu)(R\setminus\{v\}\cup\{u\}). \tag{22}$$

We now show how to prove Eq. (22). Let $\tilde{\nu}_1$ be a probability measure on $2^{[n]}$ such that $\tilde{\nu}_1(S) = \nu_1(\chi_S)$, where χ_S is as defined in (18), and where

$$\nu_1(x) \propto \exp\left(-\frac{\gamma}{2n}\left(\sum_i x_i\right)^2 + \langle h, x\rangle\right).$$

Recall from Proposition 41 that $\Pi(\tilde{\nu}_1)$ is log-concave, thus Eq. (22) is satisfied if we replace ν with $\tilde{\nu}_1$. Moreover, we consider $\Pi(\rho)$ as a probability distribution over $2^{X\sqcup Y}$ where as in Definition 47 ,X is the set of vertices, that is $[n]$, and Y is the set of dummy variables, also of size n. For $R \subseteq X\cup Y$, $\Pi(\rho)(R)$ is defined by

$$\Pi(\rho)(R) = \frac{\rho(R\cap X)}{\binom{X}{|R\cap X|}},$$

we have that

$$\Pi(\nu)(R) \propto \Pi(\tilde{\nu}_1)(R)\exp\left(\frac{1}{2}\langle r, Jr\rangle\right)$$

with $r = \chi_{R\cap X}$. So we only need to check that

$$\exp\left(\frac{1}{2}\langle r, Jr\rangle\right)\exp\left(\frac{1}{2}\langle r', Jr'\rangle\right) \le 2^{O(n)}\exp\left(\frac{1}{2}\langle s, Js\rangle\right)\exp\left(\frac{1}{2}\langle s', Js'\rangle\right) \tag{23}$$

where we define $r' = \chi_{R'\cap X}$, $s = \chi_{(R\setminus\{u\}\cup\{v\})\cap X}$, and $s' = \chi_{(R'\setminus\{v\}\cup\{u\})\cap X}$. There are two cases:

- s (s' resp.) is r (r' resp.) with coordinate i flipped. Then

$$\exp\left(\frac{1}{2}\langle r, Jr\rangle\right)\exp\left(-\frac{1}{2}\langle s, Js\rangle\right) = \exp\left(\sum_{j\neq i} J_{ij} r_i r_j\right) \leq \exp\left(\sum_{j\neq i}|J_{ij}|\right) = e^{O(\sqrt{n})}$$

since $\sum_{j\neq i}|J_{ij}| \leq \|J\|_\infty \leq \sqrt{n}\|J\|_{\mathrm{op}}$ and $\|J\|_{\mathrm{op}} < 1/2$ by assumption.

- s (s' resp.) is r (r' resp.) with coordinates i and j flipped. Then

$$\exp\left(\frac{1}{2}\langle r, Jr\rangle\right)\exp\left(-\frac{1}{2}\langle s, Js\rangle\right) = \exp\left(\sum_{k\neq i,j} r_k(J_{ki}r_i + J_{kj}r_j)\right) \leq \exp\left(\sum_{k\neq i,j}(|J_{ki}| + |J_{kj}|)\right) = e^{O(\sqrt{n})}.$$

The same inequality holds if we replace r with r' and s with s'. Thus Eq. (23) holds in both cases and we finished showing the exchange inequality for $\Pi(\nu)$.

For any distribution ν' on $2^{[n]}$, by the construction of the polarized walk we have that

$$\Pi(\nu' D^{\mathrm{pol}} U^{\mathrm{pol}}) = \Pi(\nu') D_{n\to(n-1)} U_{(n-1)\to n}.$$

Thus, for $t_1 = O(n\log n)$, by the result of [Ana+20] we have

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\nu(D^{\mathrm{pol}}U^{\mathrm{pol}})^{t_1} \parallel \nu) &= \mathcal{D}_{\mathrm{KL}}(\Pi(\nu(D^{\mathrm{pol}}U^{\mathrm{pol}})^{t_1}) \parallel \Pi(\nu)) \\ &= \mathcal{D}_{\mathrm{KL}}(\Pi(\nu')(D_{n\to(n-1)}U_{(n-1)\to n})^{t_1} \parallel \Pi(\nu)) \\ &\leq \mathrm{poly}(n)\end{aligned}$$

This combines with the entropy contraction implies that after $O(\frac{1}{1-2\|J\|_{\mathrm{op}}} n\log\frac{n}{\epsilon})$ steps of the polarized walk, we obtain a distribution $\hat{\nu}$ s.t. $d_{TV}(\hat{\nu}, \nu) \leq \epsilon$.

Implementing a step of the walk. Now we show each step of the polarized walk can be implemented in $O(d\log n)$ time. For each vertex i, we define the quantities $B_i(x) = \sum_j Q_{ij}x_j + h_i$ and $R_i(x) = \exp(B_i(x))$ which represent the effect of the other sites on site i. Note that we can compute $B_i(x)$ and $R_i(x)$ in $O(d)$ time given the list of neighbors of i. Throughout the algorithm, we store

- The current state x, which is a vector on the hypercube $\{\pm1\}^n$, and the set $x_+ = \{i \in [n] : x_i = 1\}$.
- The current value V. We maintain the invariant that $V := V(x) = \exp\left(\frac{1}{2}\langle x, Qx\rangle + \langle h, x\rangle\right)$.
- A data structure $\mathcal{D}$ storing tuples (i, R_i) at the vertices of a binary search tree keyed by i, which additionally allows the following operations:
  1. Sum(): output $\sum_i R_i$ for all (i, R_i) currently stored in the data structure.
  2. Range-Search(v, ℓ): given $\ell \geq 0$, output the minimum i in the subtree rooted at the node v such that
  $$\sum_{j<i} R_j \geq \ell.$$
  We omit v and simply write Range-Search(ℓ) if v is the root of the tree.

3. Update(i, R): sets R_i to R. If i doesn't already exist in the data structure then it inserts key i with value $R_i = R$ into the tree.

4. Delete(i): delete the pair (i, R_i) from the binary search tree.

We maintain the invariant that $i \in \mathcal{D}$ iff $x_i = -1$ and $R_j = \exp(B_j(x)) > 0$ for all $j \in \mathcal{D}$. With this invariant, $\mathcal{D}$ contains at most n nodes at any given time, so all operations on this data structure take $O(\log n)$ time by using the implementation specified in Proposition 87. It takes $O(nd \log n)$ time to initialize the algorithm at an arbitrary $x \in \{\pm\}^n$ by inserting all vertices assigned to $-$ into $\mathcal{D}$.

Now, we show that each step of the polarized walk can be implemented in $O(d \log n)$ time.

For the down step, in $O(\log n)$ time, we sample a coordinate i in x_+ with probability $1/n$ or $\perp$ with probability $\frac{n-|x_+|}{n}$. Let $T = x_+$. If the sample is a coordinate $i \in x_+$, in a total of $O(d \log n)$ time we perform the following updates:

- Update $T = T \setminus \{i\}$, compute $R_i = \exp(B_i(x))$ and insert (i, R_i) into $\mathcal{D}$. Update $V \leftarrow V/R_i^2$.
- Update $B_j(x) \leftarrow B_j(x) - 2Q_{ij}$ and $R_j \leftarrow R_j/\exp(2Q_{ij})$ for all $j \in \mathcal{N}(i)$ such that $x_j = -1$.
- Update $x = \chi_T$.

Note that we maintain the invariant $V = \exp(\frac{1}{2}\langle x, Qx\rangle + \langle h, x\rangle)$ with $x = \chi_T$, $R_j = \exp(B_j(x))$ and $j \in \mathcal{D}$ if $x_j = -1$, or equivalently, $j \notin T$.

For the up step, let L be the output of Sum(). Then:

1. With probability $\frac{n-|T|}{n}(\frac{L}{n} + \frac{n-|T|}{n})^{-1}$, set the new state x' s.t. $x'_+ = T$ and finish the step.

2. Otherwise (so with probability $\frac{L}{n}(\frac{L}{n} + \frac{n-|T|}{n})^{-1}$), do the following

   - Sample ℓ uniformly at random in the interval $[0, L]$. Let j be the output of Range-Search(ℓ). Note that j is sampled with probability

   $$\frac{R_j}{L} = \frac{\exp(B_j(\chi_T))}{\sum_{j' \notin T} \exp(B_{j'}(\chi_T))} = \frac{\nu(\chi_{T \cup \{j\}})}{\sum_{j' \notin T} \nu(\chi_{T \cup \{j'\}})}.$$

   - Remove j from $\mathcal{D}$, and update $V \leftarrow VR_j^2$.
   - Update $B_k(x) \leftarrow B_k(x) + 2Q_{jk}$ and $R_k \leftarrow R_k \exp(2Q_{jk})$ for all $k \in \mathcal{N}(j)$ such that $x_k = -1$.
   - Update $x = \chi_{T \cup \{j\}}$,

The total time for the up step is $O(|\mathcal{N}(j)| \log n) = O(d \log n)$. □

Proposition 87. *The data structure $\mathcal{D}$ as described in the proof of Theorem 86 can be implemented so that each operation takes $O(\log n)$ time, where n is an upper bound on the number of nodes stored in $\mathcal{D}$ at any given time.*

Proof. To implement $\mathcal{D}$, we will use a self-balancing binary search tree (BST) — for example, an AVL tree or red-black tree (see e.g. [Cor+22]). Each node of the tree stores the tuple (i, R_i) and is keyed by i, i.e., a transversal of the tree should return a list sorted in increasing order i of all tuples (i, R_i) in the tree. Each node also stores the sum of R_j for all j in its left subtree

$$S_i = \sum_{j \in \text{left-subtree}(i)} R_j.$$

Insertion/deletion/update for BST involves following the path from the root to a certain node, and we only need to update S_j along this path, which is of length $O(\log n)$ in a self-balancing BST. Thus, insertion/deletion/update takes worst case $O(\log n)$ time.

To implement Sum(), output the sum of S_j along each node in the path from the root to the element with the largest key i in $\mathcal{D}$. The algorithm only does $O(1)$ work at each node along a path from the root to a certain node, again taking $O(\log n)$ time.

To implement Range-Search(v, ℓ), repeat the following:

- If $\ell = S_v$, output v.
- If $\ell < S_v$: If the left-child v.left of v exists, recursively call Range-Search(v.left, ℓ). If not, output v.
- If $\ell > S_v$: If the right-child v.right of v exists, recursively call Range-Search(v.right, $\ell - S_v$). If not, output v.

The algorithm only needs to do $O(1)$ work at each node along a path from the root to a certain node, so it runs in time $O(\log n)$.

Let i be the correct output of Range-Search(v, ℓ) according to its specification, i.e., the minimum i in the subtree rooted at the node v such that

$$\sum_{j<i} R_j \geq \ell.$$

We prove by induction that this i is indeed the output of our implementation. Let $T(v)$ be the tree rooted at v and consider three cases:

- If $\ell = S_v = \sum_{j<v, j\in T(v)} R_j$ then $i = v$, thus our algorithm behaves correctly.
- If $\ell < S_v = \sum_{j<v, j\in T(v)} R_j$ then $v > i$, since i is the minimum key satisfying the above inequality, i.e., $i = \min\{i' \in T(v) \mid \ell \leq \sum_{j<i', j\in T(v)} R_j\}$, thus $i \in T(v.\text{left})$. For any $i' \in T(v.\text{left})$

  $$\sum_{j<i', j\in T(v.\text{left})} R_j = \sum_{j<i', j<v, j\in T(v)} R_j = \sum_{j<i', j\in T(v)} R_j$$

  thus $i = \min\{i' \in T(v.\text{left}) \mid \ell \leq \sum_{j<i', j\in T(v.\text{left})} R_j\}$, and thus by the inductive hypothesis it will be the output of Range-Search(v.left, ℓ).
- If $\ell > S_v = \sum_{j<v, j\in T(v)} R_j$ then $v < i$ and $i \in T(v.\text{right})$, since $\sum_{j<i, j\in T(v)} R_j \geq \ell > \sum_{j<v, j\in T(v)} R_j$. Analogously, $\sum_{j<i', j\in T(v)} R_j \geq \ell$ iff $i' \in T(v.\text{right})$. For any $i' \in T(v.\text{right})$

  $$S_v + \sum_{j<i', j\in T(v.\text{right})} R_j = \sum_{j<v, j\in T(v)} R_j + \sum_{j<i'} R_j = \sum_{j<i', j\in T(v)} R_j$$

  thus

  $$i = \min\left\{ i' \in T(v) \;\middle|\; \ell \leq \sum_{j<i', j\in T(v)} R_j \right\} = \min\left\{ i' \in T(v.\text{right}) \;\middle|\; \ell - S_r \leq \sum_{j<i', j\in T(v.\text{right})} R_j \right\},$$

  thus by the inductive hypothesis it will be the output of Range-Search(v.right, $\ell - S_v$).

□

7.4 Concentration of measure and entropy-transportation inequality

We say a Markov operator P with state space $\{\pm 1\}^n$ is ρ-local if $P(x,y) \neq 0$ only if $\|x-y\|^2 \leq \rho$. By a standard Herbst argument [HS19], if a ρ-local Markov chain with stationary distribution $\nu : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ has modified log-Sobolev constant α then ν exhibits sub-Gaussian concentration of Lipschitz function with constant $O(\alpha)$.

Definition 88. We say $\nu : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ satisfies sub-Gaussian concentration of Lipschitz functions with constant K if

$$\mathbb{P}_{S\sim\nu}\left[f(S) \geq \mathbb{E}_\nu[f(S)] + a\right] \leq \exp\left(-\frac{Ka^2}{2c^2}\right).$$

where $f : \{\pm 1\}^n \to \mathbb{R}$ is an arbitrary a c-Lipschitz functional with respect to the Hamming metric. This is equivalent to a W_1-entropy transport inequality [Van14, Theorem 4.8].

Corollary 89. *Under the assumptions of Theorem 86, the probability measure ν satisfies sub-Gaussian concentration of Lipschitz functions with constant $K = O\left(\frac{(1-2\|J\|_{\mathrm{op}})}{n}\right)$.*

7.5 Antiferromagnetic Ising model

Proposition 90. *Let $\nu : \{\pm 1\}^n \to \mathbb{R}_{\geq 0}$ be the antiferromagnetic Ising model on graph $G = G([n], E)$, i.e.,*

$$\nu(x) \propto \exp\left(-\frac{\beta}{2}\langle x, Ax\rangle + \langle h, x\rangle\right)$$

where A is the adjacency matrix of G, and $\beta \geq 0$ and $h \in \mathbb{R}^n$ are given parameters.

Let $d_{\max}, d_{\min}$ be maximum and mimimum degrees of G respectively. Consider the normalized Laplacian matrix L of G, i.e., $L = I - D^{-1/2}AD^{-1/2}$ where D is the diagonal matrix with $D_{i,i}$ be the degree of vertex i. The eigenvalues of L are $0 = \lambda_1 \leq \lambda_2 \leq \cdots \lambda_n \leq 2$. Let $\lambda(G) = \max_{i\neq 1}|1-\lambda_i|$.

If $0 \leq \beta \leq \frac{1-\delta}{4}(d_{\max} - d_{\min} + d_{\max}\lambda(G))^{-1}$ for $\delta > 0$, then Theorem 86 and Corollary 89 applies i.e.

1. *We can sample from $\hat{\nu}$ s.t. $d_{TV}(\hat{\nu}, \nu) \leq \epsilon$ by running the polarized walk for $O(\delta^{-1} n \log \frac{n}{\epsilon})$ steps. The total runtime is $O(\delta^{-1} n d_{\max} \log \frac{n}{\epsilon} \log n)$.*
2. *ν has sub-Gaussian concentration of Lipschitz functions.*

Proof. Let the eigenvalues of $D^{-1/2}AD^{-1/2} = I - L$ be $\beta_i = 1 - \lambda_i$. Note that the unit eigenvector associated with $\beta_1 = 1$ is $v_1 = \frac{1}{\sqrt{n}}\vec{1}$, thus we can write

$$D^{-1/2}AD^{-1/2} = \frac{1}{n}\vec{1}\vec{1}^\intercal + B$$

where, since B is symmetric, $\|B\|_{\mathrm{op}} = \max_{i\neq 1}|\beta_i| = \max_{i\neq 1}|1-\lambda_i| = \lambda(G)$. Thus

$$\begin{aligned} -\beta A &= -\frac{\beta}{n}(D^{1/2}\vec{1})(D^{1/2}\vec{1})^\intercal - \beta D^{1/2}BD^{1/2} \\ &= -\frac{\beta d_{\max}}{n}\vec{1}\vec{1}^\intercal + \frac{\beta}{n}(d_{\max}\vec{1}\vec{1}^\intercal - (D^{1/2}\vec{1})(D^{1/2}\vec{1})^\intercal) - \beta D^{1/2}BD^{1/2} \end{aligned}$$

Let

$$U = \frac{1}{n}(d_{\max}\vec{1}\vec{1}^\intercal - (D^{1/2}\vec{1})(D^{1/2}\vec{1})^\intercal), \qquad V = -D^{1/2}BD^{1/2}, \qquad J = \frac{1-\delta}{4}I + \beta(U+V).$$

Note that

$$u_{ij} = \frac{1}{n}(d_{\max} - D_i^{1/2} D_j^{1/2}) \le \frac{1}{n}(d_{\max} - d_{\min})$$

so $\|U\|_{\mathrm{op}} \le \|U\|_\infty = \max_i \sum_j |U_{ij}| \le (d_{\max} - d_{\min})$. Also,

$$\|V\|_{\mathrm{op}} \le \|D^{1/2}\|_{\mathrm{op}} \|B\|_{\mathrm{op}} \|D^{1/2}\|_{\mathrm{op}} \le d_{\max} \lambda(G)$$

thus

$$\|\beta(U+V)\|_{\mathrm{op}} \le \beta(\|U\|_{\mathrm{op}} + \|V\|_{\mathrm{op}}) \le |\beta|(d_{\max} - d_{\max} + d_{\max}\lambda(G)) \le \frac{1-\delta}{4}$$

and $0 \le J \le \frac{1-\delta}{2}$. Since adding multiples I to A doesn't change the distribution ν, we can write

$$\nu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle - \frac{\beta d_{\max}}{2n}\left(\sum_i x_i\right)^2\right)$$

with J as defined above, thus Theorem 86 applies. Also, since all non-diagonal entries of Q is same as that of $-A$, the maximum number of non-diagonal nonzero entries in each row of Q is bounded above by $d_{\max}$. □

We thus obtain the following results for the antiferromagnetic Ising model on random graphs.

Corollary 91. *Let ν be the antiferromagnetic Ising model with parameter β on G where G is a random d-regular graph. Suppose $0 \le \beta \le \frac{1-\delta}{8\sqrt{d-1}}$ for a constant $\delta > 0$. With probability $1 - o(1)$ over the random instance G, we can sample from $\hat{\nu}$ s.t. $d_{TV}(\hat{\nu}, \nu) \le \epsilon$ in $O(\delta^{-1} nd \log \frac{n}{\epsilon} \log n)$ time and ν has sub-Gaussian concentration of Lipschitz function.*

Proof. By Friedman's theorem [Fri03], with probability $1 - o(1)$, we have that $d_{\max}\lambda(G) = d\lambda(G) \le 2\sqrt{d-1} + o(1)$. □

Corollary 92. *Let ν be the antiferromagnetic Ising model with parameter β on G where $G = G(n,p)$ is a Erdos-Renyi random graph. Let $d = (n-1)p$ be the expected degree. Suppose $p > \frac{\log n}{n}$. There exists constant $C > 0$ s.t. if $0 \le \beta \le \frac{1}{C\sqrt{d \log n}}$, then with probability $1 - o(1)$ over the random instance G, we can sample from $\hat{\nu}$ s.t. $d_{TV}(\hat{\nu}, \nu) \le \epsilon$ in $O(nd \log \frac{n}{\epsilon} \log n)$ time and ν has sub-Gaussian concentration of Lipschitz function.*

Proof. By a standard Chernoff bound, with probability $1 - o(1)$, $d_{\max} \le d + O(\sqrt{d \log n})$ and $d_{\min} \ge d - O(\sqrt{d \log n})$. With probability $1 - o(1)$, G is connected. By [Coj07] and [HKP12, Theorem 1.1], $\lambda(G) \le O(\frac{1}{\sqrt{d}})$. Applying Proposition 90 gives the desired result. □

7.6 Application to fixed magnetization models

As a special case, our general theory applies to the fixed magnetization Ising model, i.e., the restriction of the Ising measure to the slice $\sum_i x_i = k$ of the hypercube $\{\pm 1\}^n$ for any k. See e.g. [Car+22]. This is because we can recover fixed magnetization as the limit $\gamma \to \infty$ of our more general model. These results will be applicable in particular to models on both ferromagnetic and antiferromagnetic expander graphs.

Proposition 93. *Suppose that integers k, n satisfy $-n \le k \le n$ and $k \equiv n \mod 2$. Suppose $J \succeq 0$ and $h \in \mathbb{R}^n$ are arbitrary and let μ be a probability measure on the slice $\{x \in \{\pm 1\}^n : \sum_i x_i = k\}$ with probability mass function*

$$\mu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle\right).$$

For every $\lambda \ge 0$, define the Ising model ν_λ on the hypercube $\{\pm 1\}^n$ by its probability mass function

$$\nu_\lambda(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle - \frac{\lambda}{2}\left(\sum_i x_i - k\right)^2\right).$$

Then $\nu_\lambda \to \mu$ as $\lambda \to \infty$.

Proof. The parity constraint is exactly the condition which determines whether $\sum_i x_i = k$ has a solution for $x \in \{\pm 1\}^n$. So the result follows because $\nu_\lambda(x) \to 0$ as $\lambda \to \infty$ for any $x \in \{\pm 1\}^n$ not satisfying the constraint $\sum_i x_i = k$, and because the conditional law of x given $\sum_i x_i = k$ under ν_λ is exactly μ. □

As the following corollary shows, the result we prove has consequences not just for the polarized walk but also the ordinary down-up walk. To be more precise, we can identify a vector $x \in \{\pm 1\}^n$ with the set $x_+ = \{i : x_i = +1\}$ of plus-valued entries, so the usual $D_{k\to k-1}$ down operator on sets corresponds to picking a random +1 entry of x and setting it to -1, and as usual the up operator $U_{k-1\to k}$ samples from the posterior on x given such an observation. As the proof of the following result shows, the down-up walk is simply a less lazy version of the polarized walk, and this results in it mixing faster when the number of + entries is small.

Corollary 94. *Let μ be as defined in Proposition 93. If $\|J\|_{\mathrm{op}} < 1/2$ then*

1. *The polarized down-operator D^{pol} has $\left(1 - \frac{1-2\|J\|_{\mathrm{op}}}{n}\right)$-entropy contraction with respect to μ.*
2. *The polarized walk on ν mixes within ϵ-TV distance in $O\left(\frac{1}{1-2\|J\|_{\mathrm{op}}} n \log(n/\epsilon)\right)$ steps.*
3. *Let $m = (n+k)/2 \in [0, n]$ be the total number of + spins under the fixed magnetization model. By identifying a spin assignment $x \in \{\pm 1\}^n$ with the set of vertices assigned to +, we can view μ as a distribution over $\binom{[n]}{m}$. The (ordinary) down-operator $D_{m\to(m-1)}$ has $\left(1 - \frac{1-2\|J\|_{\mathrm{op}}}{m}\right)$-entropy contraction with respect to μ.*
4. *The (ordinary) down-up walk on ν mixes within ϵ-TV distance in $O\left(\frac{1}{1-2\|J\|_{\mathrm{op}}} m \log(m/\epsilon)\right)$ steps.*
5. *The probability measure μ satisfies sub-Gaussian concentration of Lipschitz functions with constant $K = O\left(\frac{1-2\|J\|_{\mathrm{op}}}{n-|k|}\right)$.*

Proof. The first two points follow by combining Proposition 93 with Theorem 86 taking $\gamma = \lambda$. For the third point, first observe that from the output of the channel D^{pol} we can determine whether a +-spin was dropped, since the magnetization is fixed. Hence, by the chain rule for KL divergence we have

$$\mathcal{D}_{\mathrm{KL}}(\mu' D^{\mathrm{pol}} \parallel \mu D^{\mathrm{pol}}) = \frac{n+k}{2n}\mathcal{D}_{\mathrm{KL}}(\mu' D \parallel \mu D) + \frac{n-k}{2n}\mathcal{D}_{\mathrm{KL}}(\mu' \parallel \mu),$$

so using the contraction of the polarized down operator, we have

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\mu' D \parallel \mu D) &= \frac{2n}{n+k}\mathcal{D}_{\mathrm{KL}}(\mu' D^{\mathrm{pol}} \parallel \mu D^{\mathrm{pol}}) - \frac{n-k}{n+k}\mathcal{D}_{\mathrm{KL}}(\mu' \parallel \mu) \\ &\le \left(\frac{2n}{n+k}\left(1 - \frac{1-2\|J\|_{\mathrm{op}}}{n}\right) - \frac{n-k}{n+k}\right)\mathcal{D}_{\mathrm{KL}}(\mu' \parallel \mu) \\ &\le \left(1 - \frac{2(1-2\|J\|_{\mathrm{op}})}{n+k}\right)\mathcal{D}_{\mathrm{KL}}(\mu' \parallel \mu).\end{aligned}$$

The mixing time bound follows in the same way as before (using the exchange inequality).

For the last point, let m and m' be the number of $+$ and $-$ spins respectively, then $\min\{m, m'\} = \frac{n-|k|}{2}$. The down-up walk on the set of vertices assigned to $+$ ($-$ resp.) has modified log-Sobolev constant $\Omega\left(\frac{1-2\|J\|_{\mathrm{op}}}{m}\right)$ ($\Omega\left(\frac{1-2\|J\|_{\mathrm{op}}}{m'}\right)$ resp.). Both of these walks are 2-local walks, so the conclusion follows by the Herbst argument, same as in Corollary 89. □

Application: ferromagnetic Ising with fixed magnetization on expanders.

Proposition 95. *Let $\mu : \{\pm 1\}^n \to \mathbb{R}_{\ge 0}$ be the* canonical *Ising model on graph $G = G([n], E)$ at fixed magnetization k. More precisely, μ is a probability measure on the slice $\{x \in \{\pm 1\}^n : \sum_i x_i = k\}$ with probability mass function*

$$\mu(x) \propto \exp\left(-\frac{\beta}{2}\langle x, Ax\rangle + \langle h, x\rangle\right)$$

where A is the adjacency matrix of G, and β and $h \in \mathbb{R}^n$ are given parameters

Let $d_{\max}, d_{\min}$ be maximum and mimimum degrees of G respectively. Consider the normalized Laplacian matrix L of G, i.e., $L = I - D^{-1/2}AD^{-1/2}$ where D is the diagonal matrix with $D_{i,i}$ be the degree of vertex i. The eigenvalues of L are $0 = \lambda_1 \le \lambda_2 \le \cdots \lambda_n \le 2$. Let $\lambda(G) = \max_{i \ne 1}|1 - \lambda_i|$.

If $|\beta| \le \frac{1-\delta}{4}(d_{\max} - d_{\min} + d_{\max}\lambda(G))^{-1}$ for $\delta > 0$, then Corollary 94 applies i.e.,

1. *Let $m = (n+k)/2 \in [0, n]$ be the total number of $+$ spins under the fixed magnetization model. We can sample from $\hat{\mu}$ s.t. $d_{TV}(\hat{\mu}, \mu) \le \epsilon$ by running the down-up walk for $O(\delta^{-1} m \log \frac{n}{\epsilon})$ steps.*

   The down-up walk can implemented using $O(nd_{\max}\log n)$ time for preprocessing and $O(d_{\max}\log n)$ time for each step, hence the total runtime is $O(nd_{\max}\log n + \delta^{-1}md_{\max}\log n \log \frac{m}{\epsilon})$.

   When the external field is uniform, i.e., $h_i = \bar{h}\forall i$ and the graph is d-regular ($d_{\max} = d_{\min} = d$), then the preprocessing time and the time to implement each step can be reduced to $O(md\log(md))$ and $O(d\log(md) + \log\frac{1}{\epsilon})$ respectively, thus the total runtime is $O(\delta^{-1}m\log\frac{m}{\epsilon}(d\log(md)\log\frac{1}{\epsilon}))$.

2. *μ has sub-Gaussian concentration of Lipschitz functions with constant $K = O\left(\frac{\delta}{n-|k|}\right)$.*

Proof. As in proof of Proposition 90, we can rewrite the probability mass function of μ as

$$\mu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle - \frac{\beta d_{\max}}{2n}(\sum_i x_i)^2\right)$$

where $0 \preceq J \preceq \frac{1-\delta}{2}$. Since μ is supported on the slice $\sum_i x_i = k$, we can further rewrite μ as

$$\mu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle\right)$$

thus Corollary 94 applies.

For implementing the random walk, we use the algorithm in Theorem 86, which resulted in the stated runtime.

Below, we discuss how to improve the runtime for d-regular graphs with uniform external fields. We use the same notation as in the proof of Theorem 86. For a given assignment x, $B_i(x)$ only depends on the number of neighbors of i with + spin. In particular, if i has no such neighbors, then $B_i(x) = -d + \bar{h}$. Thus, instead of keeping track of all vertices with − spin as the algorithm in Theorem 86 does, we only need to keep track of those which have at least one neighbor with + spin. More precisely, we insert i with $x_i = -1$ and $\mathcal{N}(i) \cap T \neq \emptyset$ into $\mathcal{D}$, and maintain the set $\text{Free}_- = [n] \setminus T \setminus \mathcal{D}$. We can compute $F = \sum_{i \in \text{Free}_-} \exp(B_i(x)) = \exp(-d + \bar{h})|\text{Free}_-|$ and $L = \sum_{i \in \mathcal{D}} \exp(B_i(x))$ by calling Sum() on $\mathcal{D}$. Thus, we can implement the up step by uniformly sampling from Free_- with probability $\frac{F}{L+F}$ and (weighted) sampling from $\mathcal{D}$ with probability $\frac{L}{L+F}$ where the weighted sampling is implemented in the same way as in Theorem 86. The down step can be implemented same as in Theorem 86, except that we insert the sampled coordinate i into $\mathcal{D}$ or Free_- depending on the assignment of its neighbor.

Maintaining $\mathcal{D}$. The number of vertices in $\mathcal{D}$ is bounded by md, since the m vertices with + spin have at most md neighbors. At initialization, it takes $O(md)$ time to compute the number $n_i^+(x)$ of + neighbors of each − vertex i with at least one + neighbor. To do so, we loop through all + vertices j and their neighbors i and increase $n_i^+(x)$ by 1. Since $R_i(x)$ is a function of $n_i^+(x)$, i.e., $R_i(x) = \exp(B_i(x)) = \exp(2n_i^+ - d + \bar{h})$, we can build $\mathcal{D}$ by inserting $(i, R_i(x))$ with $n_i^+(x) > 0$[7] in total $O(md \log(md))$ time. Maintaining $\mathcal{D}$ in each down-up step takes $O(d \log(md))$ time.

Note also that computing the initial value $V(x) = -\exp(\frac{\beta}{2}\langle x, Ax\rangle + \langle h, x\rangle)$ can also be done in $O(md)$ time given $n_i^+(x)$ for i s.t. $n_i^+(x) > 0$. Indeed, let $S_1 = \{i : n_i^+(x) > 0\}$ and $S_2 = \{i : n_i^+(x) = 0 \text{ and } x_i = 1\}$, we can write

$$\log V(x) = -\frac{\beta}{2}\sum_i x_i B_i(x) = -\frac{\beta}{2}\left(\sum_{i \in S_1} x_i B_i(x) + (\bar{h} - d)(|S_2| - (n - |S_1| - |S_2|))\right)$$

Thus, computing $V(x)$ takes $|S_1| + |S_2| = O(md)$ time.

Maintaining Free_-. Since $|\text{Free}_-| \geq n - m(d+1)$, if $md \log(md) < n/4$ then uniformly sampling from $|\text{Free}_-|$ can be done by uniform sampling from $[n]$[8] then accept the sample if it is not in $T \cup \mathcal{D}$. To ensure that the failure probability stays below $\epsilon/2$ in $\text{poly}(md)$ steps of the down-up walk, we only need to do rejection sampling $\log \frac{md}{\epsilon}$ times. In this case, there is no need to keep an explicit data structure for Free_-, and the preprocessing time is $O(md \log(md))$. Each step of the down-up walk takes $O(d \log(md) + \log(\frac{md}{\epsilon})) = O(d \log(md) + \log(\frac{1}{\epsilon}))$ time to maintain $\mathcal{D}$ and Free_-.

If $md \log(md) > n/4$ then we can keep a data structure (i.e., a binary search tree) for Free_- that enables uniform sampling for preprocessing time of $O(|\text{Free}_-|) = O(n) = O(md \log(md))$, and the sampling time is $O(\log|\text{Free}_-|) = O(\log(md))$. Each step of the down-up walk takes $O(d \log(md))$ time.

□

[7] There are $O(md)$ such i.

[8] Each sample can be produced in constant time by hashing.

Corollary 96 (Fixed-magnetization on random-d-regular graphs and expanders)**.** *Let μ be the canonical (ferromagnetic or antiferromagnetic) Ising model with fixed magnetization k and parameter β on graph G. If G is a random-d regular graph and $\beta < \frac{1}{8\sqrt{d-1}}$, with probability $1 - o(1)$ over the random instance G, the down-up walk on μ mixes in $O_\beta(m \log \frac{m}{\epsilon})$ steps.*

Corollary 97 (Fixed-magnetization on Erdos-Renyi random graphs)**.** *Let μ be the canonical (ferromagnetic or antiferromagnetic) Ising model with fixed magnetization k and parameter β on G where $G = G(n, p)$ is an Erdos-Renyi random graph. Let $d = (n-1)p$ be the expected degree. Suppose $p > \frac{\log n}{n}$. Let $m = \frac{n+k}{2}$. There exists constant $C > 0$ s.t. if $0 \leq \beta \leq \frac{1}{C\sqrt{d \log n}}$, then with probability $1 - o(1)$ over the random instance G, the down-up walk on μ mixes in $O(m \log \frac{m}{\epsilon})$ steps.*

References

[Adh+22] Arka Adhikari, Christian Brennecke, Changji Xu, and Horng-Tzer Yau. *Spectral Gap Estimates for Mixed p-Spin Models at High Temperature*. 2022. DOI: 10.48550/ARXIV.2208.07844.

[AG22] Ahmed El Alaoui and Jason Gaitonde. "Bounds on the covariance matrix of the Sherrington-Kirkpatrick model". In: *arXiv preprint arXiv:2212.02445* (2022).

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. 118. Cambridge university press, 2010.

[AKV24] Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. "Trickle-Down in Localization Schemes and Applications". In: *Proceedings of the 56th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, June 2024.

[AL20] Vedat Levi Alev and Lap Chi Lau. "Improved analysis of higher order random walks and applications". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. 2020, pp. 1198–1211.

[ALG22] Dorna Abdolazimi, Kuikui Liu, and Shayan Oveis Gharan. "A matrix trickle-down theorem on simplicial complexes and applications to sampling colorings". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2022, pp. 161–172.

[Ali+21] 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*. STOC 2021. Virtual, Italy: Association for Computing Machinery, 2021, pp. 433–446. ISBN: 9781450380539. DOI: 10.1145/3406325.3451123.

[ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. "Spectral Independence in High-Dimensional Expanders and Applications to the Hardcore Model". In: *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, 2020.

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

[Ana+18] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. "Log-concave polynomials III: Mason's ultra-log-concavity conjecture for independent sets of matroids". In: *arXiv preprint arXiv:1811.01600* (2018).

[Ana+19a] 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: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 1–12.

[Ana+19b] 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: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. ACM, June 2019.

[Ana+20] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. "Log-Concave Polynomials IV: Exchange Properties, Tight Mixing Times, and Faster Sampling of Spanning Trees". In: *CoRR* abs/2004.07220 (2020). arXiv: 2004.07220.

[Ana+21a] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Entropic Independence I: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Distributions and High-Temperature Ising Models". In: *arXiv preprint arXiv:2106.04105* (2021).

[Ana+21b] 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". In: *arXiv preprint arXiv:2111.03247* (2021).

[Ana+21c] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, Cynthia Vinzant, and Thuy-Duong Vuong. "Log-concave polynomials IV: approximate exchange, tight mixing times, and near-optimal sampling of forests". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 408–420.

[Ana+23] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. "Universality of Spectral Independence with Applications to Fast Mixing in Spin Glasses". In: *arXiv preprint arXiv:2307.10466* (2023).

[BB08] Julius Borcea and Petter Brändén. "The Lee-Yang and Pólya-Schur programs. II. Theory of stable polynomials and applications". In: *Communications on Pure and Applied Mathematics* 62 (2008).

[BB19] Roland Bauerschmidt and Thierry Bodineau. "A very simple proof of the LSI for high temperature spin systems". In: *Journal of Functional Analysis* 276.8 (2019), pp. 2582–2588.

[BBD23] Roland Bauerschmidt, Thierry Bodineau, and Benoit Dagallier. "Kawasaki dynamics beyond the uniqueness threshold". In: 2023.

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

[Ber71] VL314399 Berezinskii. "Destruction of long-range order in one-dimensional and two-dimensional systems having a continuous symmetry group I. Classical systems". In: *Sov. Phys. JETP* 32.3 (1971), pp. 493–500.

[Bes75] Julian Besag. "Statistical analysis of non-lattice data". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 24.3 (1975), pp. 179–195.

[BGL+14] Dominique Bakry, Ivan Gentil, Michel Ledoux, et al. *Analysis and geometry of Markov diffusion operators*. Vol. 103. Springer, 2014.

[BH20] Petter Brändén and June Huh. "Lorentzian polynomials". In: *Annals of Mathematics* 192.3 (2020), pp. 821–891.

[BL02] Herm Jan Brascamp and Elliot H Lieb. "Some inequalities for Gaussian measures and the long-range order of the one-dimensional plasma". In: *Inequalities: Selecta of Elliott H. Lieb*. Springer, 2002, pp. 403–416.

[Bla+22] Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. *Identity Testing for High-Dimensional Distributions via Entropy Tensorization*. 2022. DOI: 10.48550/ARXIV.2207.09102.

[Bre+22] Christian Brennecke, Adrien Schertzer, Changji Xu, and Horng-Tzer Yau. "The Two Point Function of the SK Model without External Field at High Temperature". In: *arXiv preprint arXiv:2212.14476* (2022).

[BXY23] Christian Brennecke, Changji Xu, and Horng-Tzer Yau. "Operator Norm Bounds on the Correlation Matrix of the SK Model at High Temperature". In: *arXiv preprint arXiv:2307.12535* (2023).

[Car+22] Charlie Carlson, Ewan Davies, Alexandra Kolla, and Will Perkins. "Computational thresholds for the fixed-magnetization Ising model". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, pp. 1459–1472.

[CC09] Eric A Carlen and Dario Cordero–Erausquin. "Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities". In: *Geometric and Functional Analysis* 19.2 (2009), pp. 373–405.

[CE22] Yuansi Chen and Ronen Eldan. "Localization schemes: A framework for proving mixing bounds for Markov chains". In: *arXiv preprint arXiv:2203.04163* (2022).

[Cel22] Michael Celentano. "Sudakov-Fernique post-AMP, and a new proof of the local convexity of the TAP free energy". In: *arXiv preprint arXiv:2208.09550* (2022).

[CGM19a] Mary Cryan, Heng Guo, and Giorgos Mousa. "Modified log-Sobolev inequalities for strongly log-concave distributions". In: *arXiv preprint arXiv:1903.06081* (2019).

[CGM19b] Mary Cryan, Heng Guo, and Giorgos Mousa. "Modified log-Sobolev inequalities for strongly log-concave distributions". In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2019, pp. 1358–1370.

[Che+21] Sinho Chewi, P. Dean Gerber, Chen Lu, Thibaut Le Gouic, and Philippe Rigollet. "The query complexity of sampling from strongly log-concave distributions in one dimension". In: *Annual Conference Computational Learning Theory*. 2021.

[CLV21a] 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 (STOC)*. 2021, pp. 1537–1550.

[CLV21b] 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*. 2021, pp. 1537–1550.

[CMT14] Pietro Caputo, Georg Menz, and Prasad Tetali. *Approximate tensorization of entropy at high temperature*. 2014. DOI: 10.48550/ARXIV.1405.0608.

[Coj+22] Amin Coja-Oghlan, Philipp Loick, Balázs F Mezei, and Gregory B Sorkin. "The Ising antiferromagnet and max cut on random regular graphs". In: *SIAM Journal on Discrete Mathematics* 36.2 (2022), pp. 1306–1342.

[Coj07] Amin Coja-Oghlan. "On the Laplacian Eigenvalues of Gn,p". In: *Combinatorics, Probability and Computing* 16 (2007), pp. 923–946.

[Cor+22] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2022.

[CZS22] Xiang Cheng, Jingzhao Zhang, and Suvrit Sra. "Efficient Sampling on Riemannian Manifolds via Langevin MCMC". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 5995–6006.

[Des+19] Yash Deshpande, Andrea Montanari, Ryan O'Donnell, Tselil Schramm, and Subhabrata Sen. "The threshold for SDP-refutation of random regular NAE-3SAT". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2019, pp. 2305–2321.

[DK17] Irit Dinur and Tali Kaufman. "High dimensional expanders imply agreement expanders". In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2017, pp. 974–985.

[DLS78] Freeman J Dyson, Elliott H Lieb, and Barry Simon. "Phase Transitions in Quantum Spin Systems with Isotropic and Nonisotropic Interactions". In: *Journal of Statistical Physics* 18.4 (1978).

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

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

[Eld20] Ronen Eldan. "Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation". In: *Probability Theory and Related Fields* 176.3 (2020), pp. 737–755.

[Ell06] Richard S Ellis. *Entropy, large deviations, and statistical mechanics*. Vol. 1431. 821. Taylor & Francis, 2006.

[EM22] Ahmed El Alaoui and Andrea Montanari. "An information-theoretic view of stochastic localization". In: *IEEE Transactions on Information Theory* 68.11 (2022), pp. 7423–7426.

[EMZ20] Ronen Eldan, Dan Mikulincer, and Alex Zhai. "The CLT in high dimensions: quantitative bounds via martingale embedding". In: (2020).

[ES22] Ronen Eldan and Omer Shamir. "Log concavity and concentration of Lipschitz functions on the Boolean hypercube". In: *Journal of Functional Analysis* 282.8 (2022), p. 109392.

[Fri03] Joel Friedman. "A proof of alon's second eigenvalue conjecture". In: *Symposium on the Theory of Computing*. 2003.

[Gar73] Howard Garland. "p-Adic curvature and the cohomology of discrete subgroups of p-adic groups". In: *Annals of Mathematics* 97.3 (1973), pp. 375–423.

[GJ07] Leslie Ann Goldberg and Mark Jerrum. "The complexity of ferromagnetic Ising with local fields". In: *Combinatorics, Probability and Computing* 16.1 (2007), pp. 43–61.

[GJ19] Reza Gheissari and Aukosh Jagannath. "On the spectral gap of spherical spin glass dynamics". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 55. 2. Institut Henri Poincaré. 2019, pp. 756–776.

[GŠV16] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. "Inapproximability of the partition function for the antiferromagnetic Ising and hard-core models". In: *Combinatorics, Probability and Computing* 25.4 (2016), pp. 500–559.

[Gur09] Leonid Gurvits. "On multivariate Newton-like inequalities". In: *Advances in Combinatorial Mathematics: Proceedings of the Waterloo Workshop in Computer Algebra 2008*. Springer. 2009, pp. 61–78.

[HKP12] Christopher Hoffman, Matthew Kahle, and Elliot Paquette. "Spectral Gaps of Random Graphs and Applications". In: *International Mathematics Research Notices* (2012).

[Hop82] John J Hopfield. "Neural networks and physical systems with emergent collective computational abilities". In: *Proceedings of the national academy of sciences* 79.8 (1982), pp. 2554–2558.

[HS19] Jonathan Hermon and Justin Salez. "Modified log-Sobolev inequalities for strong-Rayleigh measures". In: *arXiv preprint arXiv:1902.02775* (2019).

[Hsu02] Elton P Hsu. *Stochastic analysis on manifolds*. 38. American Mathematical Soc., 2002.

[Hub59] John Hubbard. "Calculation of partition functions". In: *Physical Review Letters* 3.2 (1959), p. 77.

[JVV86] Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. "Random generation of combinatorial structures from a uniform distribution". In: *Theoretical computer science* 43 (1986), pp. 169–188.

[KHR22] Frederic Koehler, Alexander Heckett, and Andrej Risteski. "Statistical Efficiency of Score Matching: The View from Isoperimetry". In: *arXiv preprint arXiv:2210.00726* (2022).

[KLR22] Frederic Koehler, Holden Lee, and Andrej Risteski. "Sampling Approximately Low-Rank Ising Models: MCMC meets Variational Methods". In: *arXiv preprint arXiv:2202.08907* (2022).

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

[KO18] Tali Kaufman and Izhar Oppenheim. "High order random walks: Beyond spectral gap". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2018.

[Koe+23] Frederic Koehler, Noam Lifshitz, Dor Minzer, and Elchanan Mossel. "Influences in Mixing Measures". In: *arXiv preprint arXiv:2307.07625* (2023).

[KP21] Bo'az Klartag and Eli Putterman. "Spectral monotonicity under Gaussian convolution". In: *arXiv preprint arXiv:2107.09496* (2021).

[KS14] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*. Vol. 113. springer, 2014.

[KT18] John Michael Kosterlitz and David James Thouless. "Ordering, metastability and phase transitions in two-dimensional systems". In: *Basic Notions Of Condensed Matter Physics*. CRC Press, 2018, pp. 493–515.

[Kun23] Dmitriy Kunisky. "Optimality of Glauber dynamics for general-purpose Ising model sampling and free energy approximation". In: *arXiv preprint arXiv:2307.12581* (2023).

[LE20] Mufan Bill Li and Murat A Erdogdu. "Riemannian langevin algorithm for solving semidefinite programs". In: *arXiv preprint arXiv:2010.11176* (2020).

[Lee23] Holden Lee. "Parallelising Glauber dynamics". In: *arXiv preprint arXiv:2307.07131* (2023).

[Lit74] William A Little. "The existence of persistent states in the brain". In: *Mathematical biosciences* 19.1-2 (1974), pp. 101–120.

[LL69] Vangipuram Lakshmikantham and Srinivasa Leela. *Differential and integral inequalities: theory and applications: volume I: ordinary differential equations*. Academic press, 1969.

[LST21] Yin Tat Lee, Ruoqi Shen, and Kevin Tian. "Lower bounds on Metropolized sampling methods for well-conditioned distributions". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 18812–18824.

[LV23] Aditi Laddha and Santosh S Vempala. "Convergence of gibbs sampling: coordinate Hit-and-Run mixes fast". In: *Discrete & Computational Geometry* (2023), pp. 1–20.

[Mar13] Katalin Marton. "An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces". In: *Journal of Functional Analysis* 264.1 (2013), pp. 34–61.

[Mar99] Fabio Martinelli. "Lectures on Glauber dynamics for discrete spin models". In: *Lectures on probability theory and statistics*. Springer, 1999, pp. 93–191.

[MOS13] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. "On reverse hypercontractivity". In: *Geometric and Functional Analysis* 23.3 (2013), pp. 1062–1097.

[MP67] Vladimir Alexandrovich Marchenko and Leonid Andreevich Pastur. "Distribution of eigenvalues for some sets of random matrices". In: *Matematicheskii Sbornik* 114.4 (1967), pp. 507–536.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. Vol. 9. World Scientific Publishing Company, 1987.

[MW23] Andrea Montanari and Yuchen Wu. "Posterior sampling from the spiked models via diffusion processes". In: *arXiv preprint arXiv:2304.11449* (2023).

[MWW09] Elchanan Mossel, Dror Weitz, and Nicholas Wormald. "On the hardness of sampling independent sets beyond the tree threshold". In: *Probability Theory and Related Fields* 143.3 (2009), pp. 401–439.

[NS22] Hariharan Narayanan and Piyush Srivastava. "On the mixing time of coordinate Hit-and-Run". In: *Combinatorics, Probability and Computing* 31.2 (2022), pp. 320–332.

[Opp18] Izhar Oppenheim. "Local spectral expansion approach to high dimensional expanders part I: Descent of spectral gaps". In: *Discrete & Computational Geometry* 59.2 (2018), pp. 293–330.

[PF77] Leonid A Pastur and Alexander L Figotin. "Exactly soluble model of a spin glass". In: *Sov. J. Low Temp. Phys* 3.6 (1977), pp. 378–383.

[PW17] Yury Polyanskiy and Yihong Wu. "Strong data-processing inequalities for channels and Bayesian networks". In: *Convexity and Concentration*. Springer. 2017, pp. 211–249.

[Rei72] William Thomas Reid. *Riccati differential equations*. Vol. 86. Academic press, 1972.

[SG73] Barry Simon and Robert B Griffiths. "The (φ 4) 2 field theory as a classical Ising model". In: *Communications in Mathematical Physics* 33 (1973), pp. 145–164.

[SK75] David Sherrington and Scott Kirkpatrick. "Solvable model of a spin-glass". In: *Physical review letters* 35.26 (1975), p. 1792.

[SS12] Allan Sly and Nike Sun. "The computational hardness of counting in two-spin models on d-regular graphs". In: *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*. IEEE. 2012, pp. 361–369.

[Sta68] H Eugene Stanley. "Dependence of critical properties on dimensionality of spins". In: *Physical Review Letters* 20.12 (1968), p. 589.

[Tal10] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*. Vol. 54. Springer Science & Business Media, 2010.

[Tal11] Michel Talagrand. *Mean field models for spin glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*. Vol. 55. Springer Science & Business Media, 2011.

[Van14] Ramon Van Handel. *Probability in high dimension*. Tech. rep. PRINCETON UNIV NJ, 2014.

[Wei06] Dror Weitz. "Counting independent sets up to the tree threshold". In: *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*. 2006, pp. 140–149.

[Wu06] Liming Wu. "Poincaré and transportation inequalities for Gibbs measures under the Dobrushin uniqueness condition". In: *The Annals of Probability* 34.5 (2006), pp. 1960–1989.

[ZQM11] Zhengliang Zhang, Bin Qian, and Yutao Ma. "Uniform logarithmic Sobolev inequality for Boltzmann measures with exterior magnetic field over spheres". In: *Acta applicandae mathematicae* 116.3 (2011), pp. 305–315.

A Illustrative examples of trickle-down

A.1 Illustration: Oppenheim's trickle-down

For illustrative purposes, we show how Oppenheim's celebrated trickle-down result [Opp18] indeed follows by carefully bounding the trickle-down equation Eq. (12) applied with $T = 1$ and $t = 0$, i.e., from the equation

$$\operatorname{cov}(\nu_0) = \mathbb{E}[\operatorname{cov}(\nu_1)] + \frac{1}{k}\operatorname{cov}(\nu_0)N_0^{-1}\operatorname{cov}(\nu_0). \tag{24}$$

where $N_0 = \operatorname{diag}(\operatorname{mean}(\nu_0))$ is a diagonal matrix encoding the marginals of the measure ν_0. Also, recall that N_1 denotes a diagonal matrix encoding the marginals of the link, so it has entries $(N_1)_{ii} = \nu_1(i)$ for $i \neq a_1$ and $(N_1)_{a_1 a_1} = 0$.

The proof will be equivalent to the usual one, but we will use slightly different terminology and notation consistent with the rest of this paper. The following fact is not needed for the proof, but is a helpful reminder of the link between spectral independence and the $1 \to k \to 1$ up-down walk:

Lemma 98 ([ALO20])**.** *Suppose that ν is a probability measure on $\binom{[n]}{k}$. Then ν is C-spectrally independent iff $\lambda_2(P) \le C/k$, where $P = U_{1\to k}D_{k\to 1}$ is the (lazy version of the) $1 \to k \to 1$ up-down walk.*

See e.g. the preliminaries of [Ana+23] for a self-contained proof of the previous lemma with identical notation.

Recall from Proposition 51 that C-spectral independence is equivalent to the statement $\operatorname{cov}(\nu) \preceq Ck\Pi$ where $\operatorname{cov}(\nu)$ is the covariance matrix of the random vector 1_S for $S \sim \nu$, and $\Pi = \operatorname{diag}(\pi)$ where $\pi_i = \nu(i)/k$. The following lemma gives a slightly tighter PSD inequalilty which is yet another equivalent formulation of spectral independence.

Lemma 99. *Suppose that ν is a probability measure on $\binom{[n]}{k}$ which is C-spectrally independent, equivalently $\operatorname{cov}(\nu) \preceq Ck\Pi$. Then in fact*

$$\operatorname{cov}(\nu) \preceq Ck(\Pi - \pi\pi^\intercal).$$

Proof. First recall (by expanding the definitions) that if $P = U_{1\to k}D_{k\to 1}$ is the (lazy version of the) $1 \to k \to 1$ up-down walk, then

$$\operatorname{cov}(\nu_0) = k^2(\Pi P - \pi\pi^\intercal)$$

where $\pi_i = (1/k)\nu_0(i)$ and $\Pi = \operatorname{diag}(\pi)$ so $N_0 = k\Pi$.

Define $\lambda = C/k$, which by Lemma 98 is an upper bound on the second eigenvalue of P. Recall that P is self-adjoint with respect to the inner product $u, v \mapsto u^\intercal \Pi v$. Let $u_1 = \vec{1}, u_2, \ldots$ be the right-eigenbasis of P guaranteed by the spectral theorem (orthogonal under that inner product) with eigenvalues $\lambda_1 = 1, \lambda_2, \ldots$, and let f be arbitrary with components $f_i = (f^\intercal \Pi u_i)u_i$ so by Plancherel's theorem $f^\intercal \Pi f = \sum_i f_i^\intercal \Pi f_i$. Then for any f,

$$f^\intercal \Pi P f = \sum_i \lambda_i f^\intercal \Pi f_i = \sum_i \lambda_i f_i^\intercal \Pi f_i \le (1-\lambda) f_1^\intercal \Pi f_1 + \lambda \sum_i f_i^\intercal \Pi f_i = (1-\lambda)(f^\intercal \pi)^2 + \lambda f^\intercal \Pi f$$

which proves that $\Pi P \preceq (1-\lambda)\pi\pi^\intercal + \lambda\Pi$. Multiplying by k^2 proves the result. □

Lemma 100. *Suppose that ν is a probability measure on $\binom{[n]}{k}$ and $\lambda_2(P)$ is the second-largest eigenvalue of $P = U_{1\to k}D_{k\to 1}$, the (lazy) $1 \to k \to 1$ up-down walk. Let u_2 be the corresponding right eigenvector of P. Then*

$$\operatorname{cov}(\nu)u_2 = k^2\lambda_2(P)\Pi u_2.$$

Proof. Recall, as in the proof of Lemma 99, that $\mathrm{cov}(\nu) = k^2(\Pi P - \pi\pi^\intercal)$ where $\pi_i = (1/k)\nu_0(i)$ and $\Pi = \mathrm{diag}(\pi)$ so $N_0 = k\Pi$. Recall that the largest right eigenvector of P is the all-ones vector, and let u_2 be the second largest right eigenvector of P, which has eigenvalue $\lambda_2(P) < 1$. Recall that P is self-adjoint with respect to the inner product $u, v \mapsto u^\intercal \Pi v$ and so by spectral theorem we have $u_2^\intercal \pi = u_2^\intercal \Pi \vec{1} = 0$. From the above facts, we have that

$$\mathrm{cov}(\nu)u_2 = k^2(\Pi P - \pi\pi^\intercal)u_2 = k^2\lambda_2(P)\Pi u_2.$$

□

The following statement is Oppenheim's trickledown theorem. It may look a bit different from the usual statement, because we have stated it for a lazy version of the $1 \to k \to 1$ up-down walk, but it is directly seen to be equivalent to the "active" version — see Remark 13 of [Ana+23]. In particular, note that when $C = 1$, trickle-down preserves the property of being 1-spectrally independent inherited from the link, so trickle-down can easily be applied recursively in this case.

Theorem 101 ([Opp18])**.** *With the notation as above, suppose that almost surely*

$$\Sigma_1 \preceq CN_1,$$

and $\lambda_2(P) < 1$ where $P = U_{1\to k}D_{k\to 1}$ is the (lazy version of the) $1 \to k \to 1$ up-down walk. Then

$$\Sigma_0 \preceq \frac{C(k-2)}{k-1-C}N_0.$$

Proof. First observe from Eq. (24) and Lemma 99

$$\Sigma_0 = \mathbb{E}[\Sigma_1] + \frac{1}{k}\Sigma_0 N_0^{-1}\Sigma_0 \tag{25}$$

$$\preceq C(k-1)\,\mathbb{E}[\Pi_1 - \pi_1\pi_1^\intercal] + \frac{1}{k}\Sigma_0 N_0^{-1}\Sigma_0 \tag{26}$$

$$= C(k-1)\,\mathbb{E}[\Pi_1 - \pi\pi^\intercal] - C(k-1)(\mathbb{E}[\pi_1\pi_1^\intercal] - \pi\pi^\intercal) + \frac{1}{k}\Sigma_0 N_0^{-1}\Sigma_0 \tag{27}$$

where $\Pi_1 = \mathrm{diag}(\pi_1)$ and $(\pi_1)_i = \nu_1(i)/(k-1)$ for $i \neq a_1$, $(\pi_1)_{a_1} = 0$. (In other words π_1 is the marginal of the "link".) Note that for $S \sim \nu_0$

$$(k-1)\pi_1 = \mathbb{E}[1_{S\setminus a_1} \mid a_1] = \mathbb{E}[1_S \mid a_1] - 1_{a_1}$$

so by the law of total expectation

$$(k-1)\,\mathbb{E}[\pi_1] = \mathbb{E}[1_S] - \mathbb{E}[1_{a_1}] = (k-1)\pi,$$

i.e., $\mathbb{E}[\pi_1] = \pi$. Note that

$$\mathbb{E}\left[1_{S\setminus a_1}1_{S\setminus a_1}^\intercal\right]_{ij} = \begin{cases}(1-2/k)\,\mathbb{E}\left[1_S 1_S^\intercal\right]_{ij} & \text{if } i \neq j\\ (1-1/k)\,\mathbb{E}[1_S]_i & \text{if } i = j\end{cases}$$

so

$$\begin{aligned}\mathrm{cov}(1_{S\setminus a_1}) &= \mathbb{E}\left[1_{S\setminus a_1}1_{S\setminus a_1}^\intercal\right] - \mathbb{E}\left[1_{S\setminus a_1}\right]\mathbb{E}[1_{S\setminus a_1}^\intercal]\\ &= (1-2/k)\,\mathbb{E}[1_S 1_S^\intercal] + \Pi - (1-1/k)^2\,\mathbb{E}[1_S]\,\mathbb{E}[1_S^\intercal]\\ &= (1-2/k)\Sigma_{\nu_0} + \Pi - \pi\pi^\intercal\end{aligned}$$

so by the law of total variance and Eq. (24)

$$\begin{aligned}(k-1)^2[\mathbb{E}[\pi_1\pi_1^\intercal] - \pi\pi^\intercal] &= \operatorname{cov}(\mathbb{E}[1_{S\setminus a_1} \mid a_1]) \\ &= \operatorname{cov}(1_{S\setminus a_1}) - \mathbb{E}[\operatorname{cov}(1_{S\setminus a_1} \mid a_1)] \\ &= (1-2/k)\Sigma_0 + \Pi - \pi\pi^\intercal - \mathbb{E}[\Sigma_1] \\ &= \frac{1}{k}\Sigma_0 N_0^{-1}\Sigma_0 - \frac{2}{k}\Sigma_0 + \Pi - \pi\pi^\intercal.\end{aligned}$$

Substituting into Eq. (27) yields

$$\begin{aligned}\Sigma_0 &\le C(k-1)[\Pi - \pi\pi^\intercal] - C(k-1)(\mathbb{E}[\pi_1\pi_1^\intercal] - \pi\pi^\intercal) + \frac{1}{k}\Sigma_0 N_0^{-1}\Sigma_0 \\ &= C(k-1)[\Pi - \pi\pi^\intercal] - \frac{C}{k-1}\left(\frac{1}{k}\Sigma_0 N_0^{-1}\Sigma_0 - \frac{2}{k}\Sigma_0 + \Pi - \pi\pi^\intercal\right) + \frac{1}{k}\Sigma_0 N_0^{-1}\Sigma_0\end{aligned}$$

and rearranging gives

$$\left(1 - \frac{2C}{k(k-1)}\right)\Sigma_0 \le C(k-1-1/(k-1))[\Pi - \pi\pi^\intercal] + \frac{1}{k}(1 - C/(k-1))\Sigma_0 N_0^{-1}\Sigma_0 \tag{28}$$

Therefore by (28) and Lemma 100, we have for u_2 the right eigenvector of P corresponding to its second eigenvalue λ_2 that

$$\begin{aligned}\left(1 - \frac{2C}{k(k-1)}\right)k^2\lambda_2 u_2^\intercal \Pi u_2 &= \left(1 - \frac{2C}{k(k-1)}\right)u_2^\intercal \Sigma_0 u_2 \\ &\le C(k-1-1/(k-1))u_2^\intercal \Pi u_2 + \left(\frac{1}{k} - \frac{C}{k(k-1)}\right)u_2^\intercal \Sigma_0 N_0^{-1}\Sigma_0 u_2 \\ &= C(k-1-1/(k-1))u_2^\intercal \Pi u_2 + \left(1 - \frac{C}{(k-1)}\right)k^2\lambda_2^2 u_2^\intercal \Pi u_2\end{aligned}$$

i.e., λ_2 satisfies the quadratic inequality

$$0 \le C(k-1-1/(k-1)) - \left(1 - \frac{2C}{k(k-1)}\right)k^2\lambda_2 + \left(1 - \frac{C}{(k-1)}\right)k^2\lambda_2^2.$$

Multiplying by $k-1$ makes this

$$0 \le C(k^2-2k) - \left(k^2 - k - 2C\right)k\lambda_2 + (k-1-C)\,k^2\lambda_2^2 = (\lambda_2 - 1)\left((k-1-C)k^2\lambda_2 - C(k^2-2k)\right)$$

so using the assumption $\lambda_2 < 1$ we have that

$$\lambda_2 \le \frac{C(1-2/k)}{k-1-C} = \frac{\lambda(1-2/k)}{1-\lambda}$$

if we define $\lambda = C/(k-1)$. The conclusion follows from Lemma 98 and Proposition 51. □

A.2 Illustration: trickle-down of semi-log-concavity

Oppenheim's trickle-down shows how spectral independence/fractional log-concavity can trickle-down from top links. As another example of the trickle-down idea, we prove a simple and roughly analogous result about semi-log-concave measures. In the special case of strongly log-concave measures, the analogous fact follows from the Brascamp-Lieb theorem [BL02].

Theorem 102. *Suppose that μ is a γ-semi-log-concave measure on $\mathbb{R}^N$. Then for $T < \gamma$, the probability measure $\mathcal{G}_{-T}\mu$ defined by*

$$\frac{d\mathcal{G}_{-T}\mu}{d\mu}(x) \propto \exp(T\|x\|^2/2)$$

is $\frac{1}{\gamma - T}$ semi-log-concave.

Proof. Let $\nu_0 = \mathcal{G}_{-T}\mu$ and let ν_t be the stochastic localization process with identity driving matrix initialized at ν_0. By (10), for all $t \in [0, T]$ we have

$$\operatorname{cov}(\nu_t) = \mathbb{E}[\operatorname{cov}(\nu_T) \mid \mathcal{F}_t] + \int_t^T \mathbb{E}[\operatorname{cov}(\nu_s)^2 \mid \mathcal{F}_t] ds \preceq \frac{1}{\gamma} + \int_t^T \mathbb{E}[\operatorname{cov}(\nu_s)^2 \mid \mathcal{F}_t] ds.$$

Letting $f(t) = \sup_w \|\operatorname{cov}(\mathcal{T}_w \mathcal{G}_t \nu_0)\|_{OP}$ we therefore find that

$$f(t) \leq \frac{1}{\gamma} + \int_t^T f(s)^2 ds.$$

Recall that the solution of $dg/dt = -f(s)^2$ and $g(T) = 1/\gamma$ is $g(t) = \frac{1}{\gamma + t - T}$, so by a standard comparison argument $f(t) \leq \frac{1}{\gamma + t - T}$ and plugging in $t = 0$ gives the desired result. □

B Glauber dynamics for antiferromagnetic Ising models on low-degree expanders

In this appendix, we observe the following general result, which follows by combining key observations of the work [KLR22] (more specifically, Corollary B.2 there) with a version of the "annealing" argument from [CE22]. One of the applications of this result is $O(n \log n)$ time mixing of the Glauber dynamics for antiferromagnetic Ising models on *low-degree* expanders.

The resulting algorithm has an incomparable runtime guarantee versus the algorithm proposed and analyzed in [KLR22] based on nonconvex optimization and rejection sampling. As an advantage, the dependence on n in the runtime is nearly linear for the Glauber dynamics, whereas the algorithm from [KLR22] requires $\operatorname{poly}(n)$ time to output a single sample. As a disadvantage, the guarantee for Glauber dynamics has a doubly-exponential instead of single-exponential dependence on $\operatorname{Tr}(J_-)$. (In the application to antiferromagnetic Ising models on expanders, this means the guarantee for the Glauber dynamics would have a doubly-exponential dependence on the degree, but see Remark 105 below.)

Theorem 103. *Suppose that ν is a probability measure on the hypercube $\{\pm 1\}^n$*

$$\nu(x) \propto \exp\left(\frac{1}{2}\langle x, Jx\rangle + \langle h, x\rangle\right)$$

for some J, h such that $J \preceq 1 - 1/c$ (but J is not *required to be positive definite). Let $J = J_+ - J_-$ be the decomposition of J into its positive and negative definite parts. Then*

$$\|\mathrm{cov}(\nu)\|_{\mathrm{op}} \leq c e^{c\mathrm{Tr}(J_-)}$$

and ν satisfies ATE with constant at most

$$\exp\left(\int_0^T c e^{cs\mathrm{Tr}(J_-)} ds\right)$$

where $T = \lambda_{\max}(J) - \lambda_{\min}(J)$.

Proof. By Corollary B.2 of [KLR22], ν has a density with respect to another Ising model π where π has an interaction matrix of spectral diameter $\lambda_{\max}(J)$, and $\frac{d\nu}{d\pi} \leq e^{c\mathrm{Tr}(J_-)}$ where $J = J_+ - J_-$ is the decomposition of J into psd and negative definite components. It follows that for any function f,

$$\mathrm{Var}_\nu[f] = \mathbb{E}_\nu[(f - \mathbb{E}_\nu[f])^2] \leq \mathbb{E}_\nu[(f - \mathbb{E}_\pi[f])^2] = \mathbb{E}_\pi\left[\frac{d\nu}{d\pi}(f - \mathbb{E}_\pi[f])^2\right] \leq e^{c\mathrm{Tr}(J_-)} \mathrm{Var}_\pi[f].$$

In particular, if we consider f to be a linear function and use the bound on the covariance matrix of π from page 405 of [BL02], this proves the first conclusion.

We consider the stochastic localization process with driving matrix $C_t = J + \lambda I$ where $-\lambda$ is the smallest eigenvalue of J. By using the above argument to bound the covariance at every time t and appealing to entropic stability and the supermartingale property exactly as in the end of the proof of Theorem 54, we obtain the conclusion. □

As a consequence of this result, we get $O(n \log n)$ time mixing of the Glauber dynamics for antiferromagnetic Ising models on λ-spectral expanders with interactions of strength up to $O(1/\lambda)$, e.g., up to $O(1/\sqrt{d})$ on random graphs and other very good spectral expanders, *provided* that $d = O(1)$.

Corollary 104. *Suppose that*

$$\nu(x) \propto \exp\left(-\frac{\beta}{2}\langle x, Ax\rangle + \langle h, x\rangle\right)$$

where A is the adjacency matrix of a d-regular graph on n vertices and suppose that $\max\{|\lambda_2(A)|, |\lambda_n(A)|\} \leq \lambda$. If $\lambda\beta \leq 1 - 1/c$ for some $c > 0$, then ν satisfies ATE with constant at most $e^{ce^{O(\beta d)}}$.

Remark 105. For the above corollary, instead of using the results of [KLR22], we could use the trickledown-based result of Lemma 83 to bound the covariance matrix. This would reduce the range of β we can handle by a constant factor, but it would improve the dependence on d to be $e^{O(\beta d)}$.