Entropic Independence: Optimal Mixing of Down-Up Random Walks

Nima Anari[1], Vishesh Jain[1], Frederic Koehler[2], Huy Tuan Pham[1], and Thuy-Duong Vuong[1]

[1]Stanford University, {anari,visheshj,huypham,tdvuong}@stanford.edu
[2]Simons Institute, UC Berkeley, fkoehler@berkeley.edu

Abstract

We introduce a notion called entropic independence that is an entropic analog of spectral notions of high-dimensional expansion. Informally, entropic independence of a background distribution μ on k-sized subsets of a ground set of elements says that for any (possibly randomly chosen) set S, the relative entropy of a single element of S drawn uniformly at random carries at most $O(1/k)$ fraction of the relative entropy of S. Entropic independence is the analog of the notion of spectral independence, if one replaces variance by entropy. We use entropic independence to derive tight mixing time bounds, overcoming the lossy nature of spectral analysis of Markov chains on exponential-sized state spaces.

In our main technical result, we show a general way of deriving entropy contraction, a.k.a. modified log-Sobolev inequalities, for down-up random walks from spectral notions. We show that spectral independence of a distribution under arbitrary external fields automatically implies entropic independence. We furthermore extend our theory to the case where spectral independence does not hold under arbitrary external fields. To do this, we introduce a framework for obtaining tight mixing time bounds for Markov chains based on what we call restricted modified log-Sobolev inequalities, which guarantee entropy contraction not for all distributions, but for those in a sufficiently large neighborhood of the stationary distribution. To derive our results, we relate entropic independence to properties of polynomials: μ is entropically independent exactly when a transformed version of the generating polynomial of μ is upper bounded by its linear tangent; this property is implied by concavity of the said transformation, which was shown by prior work to be locally equivalent to spectral independence.

We apply our results to obtain (1) tight modified log-Sobolev inequalities and mixing times for multi-step down-up walks on fractionally log-concave distributions, (2) the tight mixing time of $O(n \log n)$ for Glauber dynamics on Ising models whose interaction matrix has eigenspectrum lying within an interval of length smaller than 1, improving upon the prior quadratic dependence on n, and (3) nearly-linear time $\widetilde{O}_\delta(n)$ samplers for the hardcore and Ising models on n-node graphs that have δ-relative gap to the tree-uniqueness threshold. In the last application, our bound on the running time does not depend on the maximum degree Δ of the graph, and is therefore optimal even for high-degree graphs, and in fact, is sublinear in the size of the graph for high-degree graphs.

1 Introduction

$\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ be a non-negative density on the k-subsets of the ground set $[n] = \{1, \ldots, n\}$. Such a density naturally defines a distribution on the k-subsets of $[n]$ given by

$$\mathbb{P}[S] \propto \mu(S).$$

We study a family of local Markov chains that can be used to approximately sample from such a distribution.

Definition 1 (Down-up random walks)**.** For a density $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, and an integer $\ell \leq k$, we define the $k \leftrightarrow \ell$ down-up random walk as the sequence of random sets $S_0, S_1, \ldots$ generated by the following algorithm:

for $t = 0, 1, \ldots$ **do**
  Select T_t uniformly at random from ℓ-sized subsets of S_t.
  Select S_{t+1} with probability $\propto \mu(S_{t+1})$ from supersets of size k of T_t.

This random walk is time-reversible and always has μ as its stationary distribution [see, e.g., ALO20]. The special case of $\ell = k - 1$ has received the most attention, especially in the literature on high-dimensional expanders [see, e.g., LLP17; KO18; DK17; KM16]. Recent works have established the utility of down-up random walks in capturing and analyzing widely studied Markov chains such as Glauber dynamics on graphical models or basis-exchange random walks on matroids [Ana+19; CGM20; AL20; ALO20; CLV20; Ali+21; Che+21b; Fen+21; JPV21; Liu21; Bla+21].

Each step of the down-up random walk can be efficiently implemented with oracle access to μ as long as $k - \ell = O(1)$. This is because the number of supersets of T_t is at most $n^{k-\ell} = \text{poly}(n)$, so we can enumerate over all of them in polynomial time. Even though the $k \leftrightarrow \ell$ random walk is interesting algorithmically only when $\ell = k - O(1)$, the entire range of down-up random walks is useful as an analysis tool. In fact, analyzing down-up walks where $\ell = 1$, and concluding mixing time of $k \leftrightarrow k - O(1)$ random walks, is the key technique behind most of the high-dimensional-expanders-based breakthroughs in Markov chain analysis [see, e.g., DK17; KO18; AL20].

In this work, we introduce the notion of entropic independence as a tool to establish tight bounds on the mixing time of the down-up random walks via lower bounding the modified log-Sobolev constant [BT06] (see Definition 23). Entropic independence is an entropy-based analog of spectral notions of high-dimensional expansion such as local spectral expansion [KM16; DK17; KO18] and spectral independence [ALO20]. The motivation behind considering entropy-based notions is that Markov chain mixing time analysis via spectral techniques is often lossy, by polynomially large factors on exponential-sized state spaces. On the other hand, entropy-based analysis of Markov chains can often yield tight mixing time bounds [BT06].

Our work introduces a novel technique to get *optimal* mixing time bounds using the rapidly growing literature on high-dimensional-expanders-based Markov chain analysis. Related prior works in this area fall into two categories:

Prior work on matroids and log-concave polynomials. Cryan, Guo, and Mousa [CGM20] established nearly-linear mixing time of $\tilde{O}(k)$[1] for the $k \leftrightarrow k - 1$ down-up random walk whenever μ has a log-concave generating polynomial and the walk is started from a good point. This improved upon the earlier bound of $\tilde{O}(k^2)$ established via spectral analysis by Anari, Liu, Oveis Gharan, and Vinzant [Ana+19]. Matroids are extremely good high-dimensional expanders [Ana+19]; unfortunately the techniques of Cryan, Guo, and Mousa [CGM20] appear to be limited to just matroids and matroid-related distributions, since they crucially use a specific threshold of high-dimensional expansion that can only be achieved by these distributions.

Prior work on bounded-degree graphical models. Chen, Liu, and Vigoda [CLV20] and Blanca, Caputo, Chen, Parisi, Stefankovic, and Vigoda [Bla+21] showed that under certain assumptions, spectral independence, a condition weaker than the extremely good high-dimensional expansion of matroids [ALO20], implies nearly-linear mixing time of $\tilde{O}(k)$ for down-up walks. This yielded breakthrough tight mixing time bounds for a wide range of distributions originating mostly from statistical physics. However, these results need some key assumptions, most of which do not appear to be inherently necessary, and seem to be just needed for the proof to work. The main assumption behind these works is that μ captures a graphical model on a bounded degree graph; that is μ is the joint distribution of a collection of k random variables arranged as nodes of a $O(1)$-degree graph, and that any two regions of the graph are conditionally independent of each other, conditioning on a set of nodes separating them. The bounded degree assumption together with the conditional independence assumption allows for μ to shatter into small independent pieces of size $\simeq \log k$ after conditioning on linearly many variables. A further assumption is that each random variable

[1]The notation $\tilde{O}(\cdot)$ suppresses logarithmic factors in both k and n. To keep the exposition simple in the introduction, we assume the random walks are started from a starting point with large enough of a probability mass; there always exists a point S where $\mathbb{P}_\mu[S] \geq 1/\binom{n}{k}$ and we assume our starting point approximately has this mass in the reported mixing times.

has domain of size at most $O(1)$ and that each element in the domain has marginal $\geq \Omega(1)$ according to μ and conditionings of μ. None of the assumptions, being graphical, being constant-degree, or having large marginals, appear to be inherently necessary, but are crucial to the proof.

This work. In this work we show that good high-dimensional expansion *under external fields* (see Definition 3) automatically implies entropy contraction inequalities, and in particular modified log-Sobolev inequalities for down-up random walks. Unlike mentioned prior works, we do not require extreme high-dimensional expansion, we do not assume conditional independence (being graphical), we do not require a lower bound on the marginals of μ, or even in the case of graphical models for the degrees of the graph to be $O(1)$; instead we require good high-dimensional expansion under *external fields*. This assumption is strictly weaker than the extreme high-dimensional expansion assumption of [CGM20], so our results give a proper generalization of the main result of [CGM20]. A distinguishing feature of our techniques is the ability to derive the optimal mixing time when it is not nearly-linear (see Section 1.1). Another feature of our techniques is the ability to handle lopsided distributions, where $\min\{\mu(S) \mid \mu(S) > 0\}$ is extremely small. Lopsided distributions provably cannot have a large (non-modified) log-Sobolev constant [BT06], but they can possibly have a large *modified* log-Sobolev constant. This distinction between modified and non-modified log-Sobolev inequalities is a barrier for the prior techniques of [CLV20; Bla+21], since they yield roughly the same log-Sobolev and modified log-Sobolev constants in the settings where they can be applied.

Informally, we call a background measure $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ entropically independent if for any (possibly randomly chosen) set S, the relative entropy of an element of S drawn uniformly at random carries at most $O(1/k)$ fraction of the relative entropy of S, a constant multiple of its "share of entropy." More precisely, entropic independence can be defined as entropy contraction of the $D_{k\to 1}$ operator, where $D_{k\to\ell}$ is the first part of the $k \leftrightarrow \ell$ random walk, i.e., it operates on a set $S \in \binom{[n]}{k}$ by uniformly sampling a size-ℓ subset of S. Note that $D_{k\to\ell}$ sends a distribution μ over $\binom{[n]}{k}$ to the distribution $\mu D_{k\to\ell}$ over $\binom{[n]}{\ell}$.

Definition 2 (Entropic independence)**.** A probability distribution μ on $\binom{[n]}{k}$ is said to be $(1/\alpha)$-entropically independent, for $\alpha \in (0, 1]$, if for all probability distributions ν on $\binom{[n]}{k}$,

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \parallel \mu D_{k\to 1}) \leq \frac{1}{\alpha k}\, \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

Entropic independence is a natural analog for spectral independence, another recently established notion by Anari, Liu, and Oveis Gharan [ALO20], if one replaces variance by entropy. For the special case where μ is defined via a graphical model, notions like entropic independence have been studied in prior works (although mostly as an interesting corollary, and not as the main tool to establish mixing times). See [e.g., Bla+21] for the notion of approximate subadditivity of entropy.

Spectral vs. entropic independence. In [ALO20], spectral independence is defined as an upper bound on the spectral norm of the pairwise correlation matrix of μ, or equivalently, an upper bound on the second largest eigenvalue of the simple (non-lazy) random walk on the 1-skeleton of μ, when viewing μ as a weighted high-dimensional expander [KO18; DK17].[2] The simple random walk on the 1-skeleton of μ samples from $\mu D_{k\to 1}$ by transitioning from $\{i\}$ to $\{j\}$ with probability proportional to $\mu D_{k\to 2}(\{i, j\})$. An upper bound on the second largest eigenvalue of this random walk is equivalent (up to a simple linear transformation) to an upper bound on the second largest eigenvalue of the $k \leftrightarrow 1$-down-up walk to sample from μ. One walk is simply a lazier version of the other. Standard results about the relationship between second largest eigenvalue and variance contraction then imply that variance contraction of $D_{2\to 1}$ with respect to $\mu D_{k\to 2}$ is equivalent to variance contraction of $D_{k\to 1}$ with respect to μ. However, such an

[2] In the original definition [DK17; KO18; ALO20], such a requirement is imposed for both μ and all links of μ, where the link of μ w.r.t. a set T is the distribution of $S - T$ given that S is sampled from μ conditioned on $T \subseteq S$. For the sake of clarity, and to avoid unnecessary assumptions on uniformity over links, we take the lone term "spectral independence" to only refer to the link of $T = \emptyset$, with the understanding that one usually requires spectral independence for all links; similarly, to derive mixing time bounds for down-up walks, we require entropic independence for all links.

equivalence does not hold when we replace variance with entropy: entropy contraction of $D_{2\to 1}$ with respect to $\mu D_{k\to 2}$ is a stronger assumption than entropic independence; it is easy to come up with natural distributions that do not have good $D_{2\to 1}$ entropy contraction. This introduces an inherent difficulty in establishing entropic independence. While spectral independence is about an $n \times n$ matrix, or the expansion of a simple graph on n nodes (whose edges are given by $\mu D_{k\to 2}$), there is no such compact object determining entropic independence; one has to look at all the k-sized sets and the full distribution μ.

We connect entropic independence to the geometry of the generating polynomial of the distribution μ. The multivariate generating polynomial $g_\mu \in \mathbb{R}[z_1, \ldots, z_n]$ associated to $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is given by

$$g_\mu(z_1, \ldots, z_n) := \sum_S \mu(S) \prod_{i \in S} z_i.$$

Definition 3 (External field)**.** For a distribution μ on $\binom{[n]}{k}$ and $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{>0}$, the notation $\lambda * \mu$ denotes the distribution μ tilted by external field λ, which is a distribution on $\binom{[n]}{k}$ given by

$$\mathbb{P}_{\lambda * \mu}[S] \propto \mu(S) \cdot \prod_{i \in S} \lambda_i.$$

Note that for any $(z_1, \ldots, z_n) \in \mathbb{R}^n_{\geq 0}$,

$$g_{\lambda * \mu}(z_1, \ldots, z_n) \propto g_\mu(\lambda_1 z_1, \ldots, \lambda_n z_n).$$

In Theorem 4, we show that a distribution μ is entropically independent exactly when a transformed version of the generating polynomial of μ can be upper bounded by its linear tangent, a property implied by concavity of the said transformation. We further show that this concavity is equivalent to fractional log-concavity [Ali+21], which is in turn equivalent to spectral independence under arbitrary external fields.

We recall the definition of α-fractional-log-concavity [Ali+21].

For $\alpha \in (0, 1]$, a distribution μ on $\binom{[n]}{k}$ is said to be α-fractionally log-concave (abbreviated as α-FLC) if $\log g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is concave, viewed as a function on $\mathbb{R}^n_{\geq 0}$. We note that 1-FLC is equivalent to complete/strong log-concavity [Ana+19; BH19].

Theorem 4. *Let μ be a probability distribution on $\binom{[n]}{k}$ and let $p = (p_1, \ldots, p_n) := \mu D_{k\to 1}$. Then, for $\alpha \in (0, 1]$,*

$$\begin{aligned} &\mu \text{ is } (1/\alpha)\text{-entropically independent} \\ &\iff \forall (z_1, \ldots, z_n) \in \mathbb{R}^n_{\geq 0},\ g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{1/k\alpha} \leq \sum_{i=1}^n p_i z_i. \end{aligned}$$

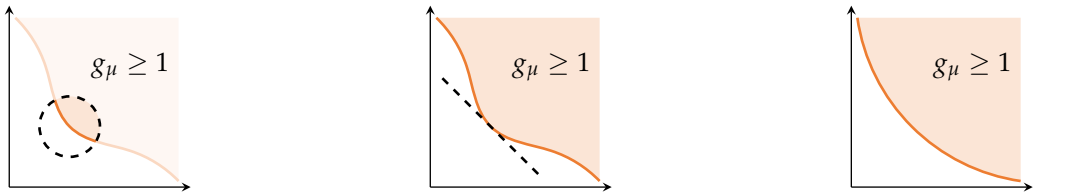
In particular, if μ is α-fractionally log-concave, then μ is $(1/\alpha)$-entropically independent. Moreover,

$$\begin{aligned} &\lambda * \mu \text{ is } (1/\alpha)\text{-entropically independent } \forall \lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{>0} \\ &\iff \mu \text{ is } \alpha\text{-fractionally log-concave.} \end{aligned}$$

Figure 1 shows a comparison of the notions in terms of the transformed generating polynomial

$$g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{1/k\alpha}.$$

Spectral independence is equivalent to *local* concavity of this function around the point $(1, \ldots, 1)$ [Ali+21]. Theorem 4 shows that entropic independence is equivalent to the linear tangent at $(1, \ldots, 1)$ upper bounding the function. Fractional log-concavity is equivalent to concavity of the entire function at all points in the positive orthant; roughly speaking, external fields allow us to replace the $(1, \ldots, 1)$ point by any other point in the positive orthant.



spectral independence | entropic independence | fractional log-concavity

Figure 1: The axes are z_1, z_2, etc. The highlighted area is a level set of the transformed generating polynomial, e.g., where $g_\mu(z_1^\alpha, \ldots, z_n^\alpha) \geq 1$. For a degree $k\alpha$-homogeneous function like $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$, log-concavity, concavity of $g_\mu^{1/k\alpha}$, and quasi-concavity, that is convexity of the level sets, are all equivalent (folklore, see Lemma 34). Notions get stronger from left to right. Spectral independence means that around the point $(1, \ldots, 1)$, the level set is locally convex. Entropic independence means that the entire level set is globally above its tangent at $(1, \ldots, 1)$. Fractional log-concavity means that the level set is globally convex.

The importance of the external fields assumption. One might a priori wish for entropy contraction or optimal mixing times from just spectral independence with no extra assumptions. In fact, Liu [Liu21] conjectured that if μ is an $O(1)$-spectrally independent distribution, the down-up walk for sampling from μ has modified log-Sobolev constant $\Omega(1/k)$. We refute this conjecture. Any α-fractionally-log-concave distribution is also $O(1/\alpha)$-spectrally independent [Ali+21, Remark 68]. However, there are examples of $\Omega(1)$-fractionally-log-concave distributions for which the down-up walk is not even irreducible — see Remark 6. Even ignoring the ergodicity issue of the walk, it is well-known that the spectral gap of a Markov chain does not *automatically* translate into a modified log-Sobolev inequality [see, e.g., BT06]. A classical example is the random walk on a constant-degree expander graph. If G is an n-node constant degree expander, then the random walk on G has $\Omega(1)$ spectral gap, but its mixing time is $\simeq \log(n)$; a constant factor entropy contraction of this random walk would imply a mixing time of $\simeq \log\log(n)$ which is clearly wrong. One can view (the lazy version of) this random walk as a special case of down-up random walks. Let $\mu : \binom{[n]}{2} \to \mathbb{R}_{\geq 0}$ be the uniform distribution on the edges of the expander. Then the $2 \leftrightarrow 1$ down-up walk is the same as the lazy random walk on G itself. This shows that even for down-up random walks, one cannot obtain entropic independence from spectral independence with no extra assumptions.

As mentioned above, for general spectrally independent distributions, the $k \leftrightarrow k-1$ down-up walk might not even be ergodic. We can fix the lack of ergodicity in the $k \leftrightarrow k-1$ down-up walk by considering the more general $k \leftrightarrow \ell$ down-up walk for smaller values of ℓ. For this more general walk, we establish a modified log-Sobolev inequality for all fractionally log-concave distributions via entropic independence. More precisely, in Theorem 5, we show that for α-fractionally-log-concave $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$, the $k \leftrightarrow (k - \lceil 1/\alpha \rceil)$ down-up walk [AL20; Ali+21], has modified log-Sobolev constant $\Omega(k^{-1/\alpha})$. The dependence on α is optimal — again, see Remark 6.

Theorem 5 is a natural generalization of a recent result of Cryan, Guo, and Mousa [CGM20, Theorem 1], which shows that the modified log-Sobolev constant of the $k \leftrightarrow (k-1)$-down-up walk for sampling from a log-concave (i.e., $\alpha = 1$) distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is at least $\Omega(\frac{1}{k})$. We remark that our proof for Theorem 5 is shorter and less "mysterious" than that of [CGM20, Theorem 1].

Theorem 5. *Suppose $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ is α-fractionally log-concave, or more generally $(1/\alpha)$-entropically independent for all links. Let $\ell \leq k - \lceil 1/\alpha \rceil$. For all probability distributions ν on $\binom{[n]}{k}$,*

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to\ell} \parallel \mu D_{k\to\ell}) \leq (1-\kappa)\, \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu).$$

Consequently, the $k \leftrightarrow \ell$ down-up walk w.r.t. μ has modified log-Sobolev constant $\geq \Omega(\kappa)$ where Ω hides an absolute constant and

$$\kappa = \frac{(k+1-\ell-1/\alpha)^{1/\alpha - \lceil 1/\alpha \rceil} \prod_{i=0}^{\lceil 1/\alpha \rceil - 1} (k-\ell-i)}{(k+1)^{1/\alpha}}.$$

For $1/\alpha \in \mathbb{Z}$ we can obtain a simpler looking alternative bound of

$$\kappa = \binom{k-\ell}{1/\alpha} \Big/ \binom{k}{1/\alpha}.$$

Remark 6. Theorem 5 is tight up to a constant. For $\alpha = 1$, it is easy to come up with a log-concave distribution μ for which the one-step down-up walk has modified log-Sobolev constant $\Theta(\frac{1}{k})$, e.g., μ uniformly supported on $\{[k], [k-1] \cup \{k+1\}\}$. For $\alpha = 1/r$ with $r \in \mathbb{N}_{\geq 1}$, we can consider the following example. Let μ be the uniform distribution on $\binom{[n]}{k'}$, and we will define a corresponding r-fold distribution $\mu^{(r)}$ supported on

$$\{S \times [r] \mid S \in \mathrm{supp}(\mu)\} \subseteq \binom{[n] \times [r]}{k'r}.$$

The probability mass function on its support is given by

$$\mu^{(r)}(S \times [r]) = \mu(S),$$

where $S \times [r]$ is the set $\{(a, i) \mid a \in S, i \in [r]\}$. Since the generating polynomial of $\mu^{(r)}$ is determined by that of μ, which is 1-FLC, we can check that this distribution is indeed $\alpha = 1/r$ fractionally log-concave.

Let $k = k'r$. Clearly, the $k \leftrightarrow (k - r')$-down-up walk w.r.t $\mu^{(r)}$ with $r' < r$ is not even irreducible, because the down-up walk always stays at the same place. On the other hand, by comparing the $k \leftrightarrow k - r$-down-up walk w.r.t $\mu^{(r)}$ with the $k' \leftrightarrow (k'-1)$-down-up walk w.r.t μ, we can show that the former has modified log-Sobolev constant $\Theta\left(1/\binom{k'r}{r}\right) = \Theta\left(\frac{r!}{k'^{1/\alpha}}\right)$.

1.1 Application: Fractionally Log-Concave Polynomials

The most straightforward application of our techniques is to establish modified log-Sobolev inequalities and tight mixing times for distributions that are spectrally independent under arbitrary external fields, a.k.a. fractionally log-concave distributions [Ali+21]. In Sections 1.2 and 1.3, we demonstrate applications beyond this setting, where one has to combine our techniques with others to establish tight mixing time bounds.

Theorem 5 is a generalization of the main results of Cryan, Guo, and Mousa [CGM20], so as a special case we recover the tight mixing time and MLSI constants established previously for all distributions with a log-concave generating polynomial [see Ana+19, for examples]. Here we highlight two important examples of fractionally log-concave distributions that go beyond simple log-concavity. We obtain, for the first time, tight mixing time bounds and MLSI constants for these examples; for further examples of fractionally log-concave distributions see [Ali+21]. We emphasize that in both examples, the tight mixing time is near-quadratic and notably not near-linear. As a result, none of the previous high-dimensional-expanders-based frameworks could obtain the tight mixing time in these examples.

Definition 7 (Monomer-Dimer Systems)**.** Suppose that a graph $G = (V, E)$ is given together with node weights $z : V \to \mathbb{R}_{\geq 0}$ and edge weights $w : E \to \mathbb{R}_{\geq 0}$. Then the monomer-dimer system is the distribution on matchings M of the graph where

$$\mathbb{P}[\text{matching } M] \propto \prod_{e \in M} w(e) \cdot \prod_{v \text{ not matched by } M} z(v).$$

For a matching M, the edges in the matching are called dimers, and the nodes outside of the matching are called monomers. For an arbitrary monomer-dimer system, we can define a measure $\mu : \binom{V \times \{0,1\}}{|V|} \to \mathbb{R}_{\geq 0}$ capturing the distribution of monomers

$$\mu(S) = \begin{cases} 0 & \exists v \in V : |S \cap \{(v,0),(v,1)\}| \neq 1, \\ \sum \{\mathrm{weight}(M) \mid \text{monomers of } M = S \cap (V \times \{1\})\} & \text{otherwise.} \end{cases}$$

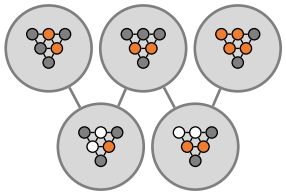


Figure 2: 2-site Glauber dynamics on monomers. A configuration consists of an assignment of binary labels indicating monomer/non-monomer to the vertices. In each step, two uniformly randomly picked vertices have their labels resampled conditioned on all other labels, with probabilities dictated by the monomer distribution.

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix}$$

Figure 3: Example of a matrix L with quadratic mixing time for down-up walks. Note that $L + L^\intercal = 0 \succeq 0$. L is block-diagonal with 2×2 blocks. The nonsymmetric determinantal point process selects independently and uniformly at random between including/excluding both elements $\{2i, 2i+1\}$ for each i.

In [Ali+21], the authors showed that the distribution of monomers μ is $\alpha = 1/2$-fractionally-log-concave and proceeded to use the $|V| \leftrightarrow |V| - 2$ down-up random walk, a.k.a. the 2-site Glauber dynamics (see Fig. 2), to show polynomial-time sampling algorithms for monomer-dimer systems on planar graphs. Alimohammadi, Anari, Shiragur, and Vuong [Ali+21] showed a spectral gap of $\Omega(1/|V|^2)$, and as a result a mixing time of $\tilde{O}(|V|^3)$ assuming a good starting point, for this random walk. As a direct corollary of Theorem 5, we obtain the following result:

Corollary 8. *The 2-site Glauber dynamics on the monomer distribution of a k vertex graph has MLSI constant $\Omega(1/k^2)$. As a result it mixes in time $\tilde{O}(k^2)$, assuming the walk is started from a configuration with probability mass $\geq 1/2^{\mathrm{poly}(k)}$.*

Note that the requirement on the starting point is fairly weak (and can even be weakened further to $1/2^{2^{\mathrm{poly}\log(k)}}$). In particular, if we find the matching M that has the maximum monomer-dimer weight using a maximum weighted matching algorithm, and start the random walk from monomers of M, that automatically satisfies the initial condition.

It is not hard to see that Corollary 8 is tight. Consider the case where the graph G itself is a perfect matching. Then, each step of the 2-site Glauber dynamics has only a $1/k$ chance of picking two endpoints of the same edge; if that does not happen, the resampling of the two vertices does not change anything and a turn is "wasted." It is also not hard to see that roughly speaking all of the $\Omega(k)$ pairs of endpoints of edges need to be resampled once before mixing. So clearly, the mixing time of this chain is $\tilde{\Omega}(k^2)$.

Our next application involves nonsymmetric determinantal point processes, a generalization of determinantal point processes which have found many uses in machine learning, recommender systems, and randomized linear algebra [see, e.g., Gar+19].

Definition 9 (Nonsymmetric Determinantal Point Process)**.** Suppose that $L \in \mathbb{R}^{k \times k}$ is such that $L + L^\intercal \succeq 0$. The distribution on subsets T of $\{1, \ldots, k\}$, giving a weight of $\det(L_{T,T})$ to each set, is called the (nonsymmetric) determinantal point process. We view this distribution as $\mu : \binom{[k] \times \{0,1\}}{k} \to \mathbb{R}_{\geq 0}$, where

$$\mu(S) = \begin{cases} 0 & \exists i : |S \cap \{(i,0),(i,1)\}| \neq 1, \\ \det(L_{T,T}) & \text{otherwise, where } T = \{i \mid (i,1) \in S\}. \end{cases}$$

Alimohammadi, Anari, Shiragur, and Vuong [Ali+21] showed that the above distribution is also $\alpha = 1/2$-fractionally log-concave, and as a result established a spectral gap of $\Omega(1/k^2)$ and a mixing time of $\tilde{O}(k^3)$ with a good choice of the starting point, for the $k \leftrightarrow k - 2$ down-up random walk.

Corollary 10. *The $k \leftrightarrow k - 2$ down-up random walk on nonsymmetric determinantal point processes has MLSI constant $\Omega(1/k^2)$. As a result, it mixes in time $\tilde{O}(k^2)$, assuming the walk is started from a set with probability mass $\geq 1/2^{\mathrm{poly}(k)}$.*

Again, finding a starting point satisfying the initial condition is not difficult; for example, it can be achieved by a simple local search [AV21]. Once again, it is not difficult to see that this quadratic mixing time is tight. The matrix L realizing a mixing time of $\tilde{\Omega}(k^2)$ can be seen in Fig. 3. The distribution defined by this particular L in Fig. 3 is the same as the one described by the tight example of monomer distributions; we leave this as an exercise to the reader.

Remark 11. Both examples mentioned above stem from the so-called Hurwitz-stable polynomials [see Ali+21]. Theorem 5 can be applied more generally to get tight mixing time bounds for arbitrary fractionally log-concave polynomials, and as a special case, the so-called sector-stable polynomials [Ali+21]. However, there currently seems to be a loss in the analysis of Alimohammadi, Anari, Shiragur, and Vuong [Ali+21] when going from sector-stability to fractional log-concavity for general sector-stable polynomials (this loss is avoided in the case of Hurwitz-stability though). As a result, while we get *improved* mixing time bounds in other examples of fractionally log-concave polynomials, the bounds do not always seem to be tight. If the analysis of [Ali+21] is improved in future works, and one can establish that homogeneous α-sector-stable polynomials are indeed α-fractionally log-concave, then combined with our results we would get a tight mixing time bound for all sector-stable polynomials.

1.2 Application: High-Temperature Ising Models

An Ising model is a probability distribution on the discrete hypercube $\{\pm 1\}^n$ given by

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

where $J \in \mathbb{R}^{n\times n}$ is a symmetric matrix (known as the interaction matrix) and $h \in \mathbb{R}^n$ is known as the external field. These models are of fundamental importance in statistical physics and other areas; see e.g. [MPV87; Lau96; MM09; Tal10].

The Glauber dynamics (or Gibbs sampler) is a simple and very popular discrete time Markov chain to sample from $\mu_{J,h}$. Its transitions may be described as follows: at each step, given the current state σ, a coordinate i is chosen uniformly at random from $[n]$ and its value is resampled from the conditional distribution $\mu_{J,h}(\cdot \mid \sigma_{-i})$; it can also be viewed as a down-up walk on a homogenized version of μ, see Definition 26. The study of the Glauber dynamics for the Ising model is a classical topic with numerous results: see e.g. [Mar99; LP17] for an introduction. Rapid mixing of the Glauber chain is interconnected with other structural properties of the Gibbs measure. In particular, it is well-known that the Dobrushin uniqueness condition – $\|J\|_{\infty\to\infty} < 1$ (equivalently, $\max_{i\in[n]} \sum_j |J_{ij}| < 1$) – implies that the Glauber dynamics mixes in time $O(n \log n)$.

Although Dobrushin's condition is a tight condition for rapid mixing in certain cases (e.g. the Curie-Weiss model, which is the Ising model on the complete graph), there are many interesting cases where Dobrushin's condition is extremely restrictive. An important example is the famous Sherrington-Kirkpatrick (SK) model from spin glass theory [MPV87; Tal10], where $J_{ij} \sim N(0, \beta^2/n)$ is a GOE random matrix. In this case, Dobrushin's condition implies rapid mixing only if the inverse temperature β is $O(1/\sqrt{n})$, even though it is widely believed that the Glauber dynamics mixes rapidly for all $\beta < c$ for an absolute constant $c > 0$, even conjecturally with $c = 1$ [Var].

Polynomial time mixing in the SK model was recently established with $c = 1/4$ in recent work of Eldan, Koehler, and Zeitouni [EKZ20] as a consequence of the following general result: if the interaction matrix J is positive semidefinite, then the spectral gap of the Glauber dynamics for the Ising model is at least $\frac{1-\|J\|_{\mathrm{OP}}}{n}$, where $\|J\|_{\mathrm{OP}}$ denotes the $\ell^2 \to \ell^2$ operator norm, which, since J is symmetric, coincides with the largest eigenvalue of J. The interaction matrix J can always be assumed to be positive semidefinite without loss of generality, because adding a multiple of the identity matrix to J does not change the measure $\mu_{J,h}$. By the well-known relationship between the spectral gap and the mixing time, the result of [EKZ20] implies that the mixing time of the Glauber dynamics is $O\left(\frac{n}{1-\|J\|_{\mathrm{OP}}} \cdot (n + \|h\|_1)\right)$, i.e. quadratic in n. However, it seemed plausible that the true mixing time was faster than this.

Here, as an application of our theory, we establish an optimal $O(n \log n)$ bound on the mixing time and also show that the modified log-Sobolev inequality holds:

Theorem 12. *Let $\mu_{J,h}$ denote an Ising measure on $\{\pm 1\}^n$ with $0 \preceq J \preceq \|J\|_{\mathrm{OP}} I$ and let P denote the transition matrix of the the Glauber dynamics. Then,*

(a) The modified log-Sobolev constant $\rho_0(P)$ of P satisfies $\rho_0(P) \geq \frac{1-\|J\|_{\mathrm{OP}}}{n}$.

(b) The Glauber dynamics mixes in time $O\left(\frac{n \log n}{1-\|J\|_{\mathrm{OP}}}\right)$.

The MLSI is used in the proof of mixing, but it also has many other useful direct consequences, such as concentration of measure and reverse hypercontractive estimates [Van14; MOS13].

Theorem 12 can be applied to a number of other models of interest. For example, a d-regular version of the diluted SK model can be formed by taking a random d-regular graph and assigning edge weights i.i.d. from $Uni(\pm\beta)$. Combining the above result with a version of Friedman's theorem (see [EKZ20] for details and references) proves the optimal $O(n \log(n))$ mixing time of Glauber dynamics for all $\beta < \frac{1}{4\sqrt{d-1}}$, whereas Dobrushin's uniqueness condition, or the more precise tree uniqueness criterion (see e.g. [SST14]), holds only when $\beta = O(1/d)$. The reason for the discrepancy is that for these spin glass models, the uniqueness threshold is not the relevant phase transition on the infinite d-ary tree. Instead, the properly analogous phase transition concerns the tree with spin-glass boundary conditions [Cha+86; PP10] or the purity of the limiting Gibbs measure with free boundary conditions [BRZ95; Iof96; Eva+00], otherwise known as the *reconstruction threshold*, which is well past the uniqueness threshold.

Comparison to previous work. Although sampling from Ising models is a classical topic, only recently in the breakthrough work of Bauerschmidt and Bodineau [BB19] was a polynomial time sampling result established for the SK model with constant β. Under the same condition $0 \preceq J \prec I$ as above, they showed how to draw a sample from the Ising model by sampling from a related log-concave distribution in $\mathbb{R}^n$ and applying an additional rounding step. They also proved a version of the log-Sobolev inequality, but their version only implies $e^{O(\sqrt{n})}$ time mixing bounds for the Glauber dynamics in the SK model — see discussion in [EKZ20].

Later, in the work [EKZ20] it was proved that the Glauber dynamics indeed mix in polynomial time. Their result established a reduction for proving functional inequalities to the case where J is rank one, and the $O(n^2)$ mixing time guarantee was established using the Poincaré inequality of [Hay06; Wu06]. However, there was no analogous way to establish the MLSI for the class of rank-one Ising models based on existing results. For example, directly applying a state of the art result such as [Mar19] gives an MLSI with constant $e^{-\Omega(\sqrt{n})}$. The issue is that in these models, like the SK model itself, the conditional marginals can be very tiny and existing methods are unable to handle this efficiently. In contrast, our approach based on entropic independence requires no assumption on boundedness of the conditional marginals and enables us to prove the MLSI.

1.3 Application: Mixing up to the Uniqueness Threshold

Our final collection of applications are to the Ising and hardcore models in the so-called tree uniqueness regime. In this subsection, for consistency with existing literature, we will slightly change our notation compared to Section 1.2 and define an Ising model on a graph $G = (V, E)$ to be the probability measure over spins $\sigma \in \{\pm 1\}^V$ with probability mass function

$$\mu(\sigma) \propto \lambda^{|\{i|\sigma_i=+1\}|} \prod_{\{i,j\}\in E} \beta^{\mathbb{1}(\sigma_i=\sigma_j)}$$

parameterized by *external field* λ and *edge activity* β, where the parameter β controls whether the spins like to align with ($\beta > 1$, *ferromagnetic*) or opposite to ($\beta < 1$, *antiferromagnetic*) their neighbors on the graph. In particular, the Ising model with external field λ and edge activity β on a graph G can be recovered as the Ising model in Section 1.2 with interaction matrix $J = \frac{\log \beta}{2} A_G$ and external field $h = \frac{\log \lambda}{2}$, where A_G is the

adjacency matrix of G. Similarly, the hardcore model is the probability measure on *independent sets* $\sigma \subset 2^V$ given by

$$\mu(\sigma) \propto \lambda^{|\sigma|},$$

i.e., weighted by the size of the independent set, where the parameter λ is conventionally referred to as the *fugacity*. The study of these two models has been closely linked; in some cases the hardcore model can even be recovered as a limit of the Ising model with strongly antiferromagnetic interactions ($\beta \to 0, \lambda \to 0$).

There has been an intense interest in understanding the sharp thresholds for mixing in the hardcore and Ising models on the class of graphs of maximum degree Δ based on connections to the *uniqueness threshold* on the infinite Δ-regular tree. After a long line of work including [DFJ02; Wei06; MWW09; Sly10; SS12; MS13; SST14; GV16; CLV20; Che+21a; ALO20] we know that in the particular case of the hardcore model, sampling is computationally hard above the uniqueness threshold on graphs of degree at most Δ [Sly10], and below the uniqueness threshold sampling can be done in $\tilde{O}(n^2)$ steps of the Glauber dynamics [Che+21a] or $\widetilde{O}(C_\Delta n)$ steps if the maximum degree Δ is fixed [CLV20; Bla+21]; C_Δ is an exponentially large function of the maximum degree Δ. The picture is similar for the Ising model, with C_Δ being instead a polynomially large function of Δ whose exponent depends on the gap to the uniqueness threshold [CLV20; Bla+21]. Nevertheless, it has been generally expected that the mixing time of Glauber dynamics is always $O(n \log n)$ within the uniqueness regime, regardless of the degree of the graph. Chen, Liu, and Vigoda [CLV20] raised the challenge of proving even a weaker bound with near-linear dependence on n and polynomial dependence on Δ, i.e., $\widetilde{O}(n \cdot \mathrm{poly}(\Delta))$, on the mixing time of Glauber dynamics for the tree-unique hardcore model. Chen, Feng, Yin, and Zhang [Che+21a] also raised the open problem of proving $\widetilde{O}(n)$ mixing time in the high-degree setting.

Optimal mixing of the (balanced) Glauber dynamics. For sampling from the hardcore model, we consider a variant of Glauber dynamics which we call the *balanced Glauber dynamics*. In this variation, the update site is chosen in a slightly non-uniform fashion, effectively introducing a small number of additional "balancing" updates into the usual Glauber chain. We explain the motivation for this small modification in Section 1.4 below.

Theorem 13. *Suppose μ is the δ-unique hardcore model on $G = (V, E)$ with $|V| = n$. The balanced Glauber dynamics with $O_\delta(n \log n)$ many steps approximately samples from μ.*

Theorem 14. *Suppose μ is the δ-unique Ising model on $G = (V, E)$ with $|V| = n$. The (standard) Glauber dynamics with $O_\delta(n \log n)$ many steps approximately samples from μ.*

Remark 15. The uniqueness region defined for the Ising model has a strange peculiarity: Theorem 14 in its exact form, that is nearly-linear mixing for all degrees, can actually be derived without appealing to the techniques of this work or even the prior work of Chen, Feng, Yin, and Zhang [Che+21a] who studied spectral gap for high-degree regimes. The reason is that δ-unique Ising models with high enough degree $\Delta > \Delta_0(\delta)$ satisfy the Dobrushin uniqueness condition! And complementing that, all small degrees $\Delta = O(1)$ are covered by the earlier work of Chen, Liu, and Vigoda [CLV20].

To see why high-degree cases fall under Dobrushin's regime, note that each entry of the Dobrushin influence matrix (see [Hay06] for definition) can be bounded by $|\log(\beta)|/2$. Be aware that the notation β is not consistent between our work and [Hay06]; one can translate $\beta \mapsto \log(\beta)/2$ to go from our notation to that of [Hay06]. This means the norm of the Dobrushin influence matrix is bounded by $\Delta \cdot |\log(\beta)|/2$ which is asymptotically

$$\frac{\Delta}{2} \cdot \left(\frac{2 - \Theta(\delta)}{\Delta} + O(1/\Delta^2) \right) = 1 - \Theta(\delta) + O(1/\Delta).$$

Fixing δ, for large enough Δ, this norm gets smaller than a constant < 1, which entails Dobrushin uniqueness and hence nearly-linear mixing time [Hay06; LP17].

We remark that "asymptotic Dorbushin uniqueness" does not happen for the hardcore model or slight variants of the tree-unique Ising model (see Remark 16 for details).

Remark 16. Note that in the limit $\Delta \to \infty$, the endpoints of the uniqueness region for the Ising model are essentially of the form $\beta = 1 \pm 2/\Delta$. For antiferromagnetic Ising models, even if β is outside of this region (e.g. $\beta \approx 1 - \alpha/\Delta$ for $\alpha > 2$), there is a critical external field $\lambda_c \in (0,1)$ below which the Ising model is in the tree uniqueness region [SST14]: we also establish an analogous result covering this two-dimensional uniqueness region in (β, λ) space using the balanced Glauber dynamics. The sublinear time sampling result (Theorem 18) and concentration result (Theorem 21) below also extend to this setting.

Sublinear time sampling algorithms. Supposing that the graph G is represented using the standard data structure of adjacency arrays, i.e., each vertex has an array of neighbors so that sampling a random neighbor of a vertex can be performed in $O(1)$ time. We show how to sample from both the hardcore model and the Ising model in runtime nearly linear in the output length n, and so in sublinear time for graphs of at least polylogarithmic average degree.

Theorem 17. *Suppose μ is the δ-unique hardcore model on $G = (V, E)$ with $|V| = n$, and G is represented by adjacency arrays. Then, there is a randomized algorithm to approximately sample from μ which can be implemented in expected time $O_\delta(n \log^2(n))$.*

Theorem 18. *Suppose μ is the δ-unique Ising model on $G = (V, E)$ with $|V| = n$ and G is represented by adjacency arrays. Then a step of the Glauber dynamics can be implemented by a randomized algorithm with expected running time $O(1)$. Combined with Theorem 14, this implies that approximate sampling can be performed in expected runtime $O_\delta(n \log n)$.*

Sharp concentration of measure and transport-entropy inequalities. Using our restricted modified log-Sobolev inequalities, we show via the Herbst argument that sub-Gaussian concentration bounds hold for all Lipschitz functions in both the hardcore and Ising model in the uniqueness region. By the celebrated result of Bobkov and Götze [BG99], concentration of Lipschitz functions is equivalent to a W_1 transport-entropy inequality, i.e. $W_1(\nu, \mu)^2 \le C\, \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$ for all measures ν where W_1 denotes the Wasserstein-1 distance with the Hamming metric.

Theorem 19. *Suppose μ is the δ-unique hardcore model on a graph with n vertices, and let f be so that $|f(\sigma_+) - f(\sigma_-)| \le \kappa$ for all adjacent states (σ_-, σ_+), i.e. f is κ-Lipschitz with respect to the Hamming metric. For all $t \ge 0$, we have for some $c = c(\delta) > 0$ that*

$$\mathbb{P}_\mu[f - \mathbb{E}_\mu[f] > t] \le e^{-ct^2/\kappa^2 n}.$$

Remark 20. In the hardcore model with small fugacity, sites are much more likely to be unoccupied than occupied, which can lead to even better concentration. To reflect this, we establish (see the full version [Ana+21b, Proposition 71]) a more precise two-level Bernstein-type inequality for monotone functionals, such as the number of occupied sites in the hardcore model.

Theorem 21. *Suppose μ is the δ-unique Ising model on a graph with n vertices, and let f be so that $|f(\sigma_+) - f(\sigma_-)| \le \kappa$ for all adjacent states (σ_-, σ_+), i.e. f is κ-Lipschitz with respect to the Hamming metric. For all $t \ge 0$ we have*

$$\mathbb{P}_\mu[f - \mathbb{E}_\mu[f] > t] \le e^{-ct^2/\kappa^2 n}$$

for some $c = c(\delta) > 0$.

1.4 Techniques

First, we discuss the idea behind Theorem 4, which in particular establishes entropic independence of the probability distribution μ given fractional log-concavity of the generating polynomial g_μ. Recall that $(1/\alpha)$-entropic independence holds for μ if the inequality

$$\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \parallel \mu D_{k\to 1}) \le \frac{1}{\alpha k}\, \mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu)$$

is true for all probability measures ν. The connection with the generating polynomial appears when we fix the left hand side, or more precisely fix the marginal $q = (q_1, \ldots, q_n) := \nu D_{k\to 1}$, and ask for the *worst-case*

choice of ν given this constraint: the measure ν minimizing the rhs. This is a minimum relative entropy problem, so based on convex duality (Lemma 35, cf. [SV14]), we get a formula in terms of "dual" variables $\log z_1, \ldots, \log z_n$ corresponding to the constraints; concretely, we have

$$\inf \{\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \mid \nu D_{k\to 1} = q\} = -\log \left(\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1,\ldots,z_n)}{z_1^{kq_1} \cdots z_n^{kq_n}} \right).$$

The factor of k in the exponent of each z_i appears because the down operator $D_{k\to 1}$ has a $1/k$ chance of picking any particular element of its input set S. Note that the rhs can now be lower bounded by making any choice of the variables $z_1, \ldots, z_n > 0$. By choosing z related to q and $p := \mu D_{k\to 1}$, we can then show that the right hand side can be lower bounded in terms of $\mathcal{D}_{\mathrm{KL}}(q \parallel p) = \mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1} \parallel \mu D_{k\to 1})$ if an appropriate convexity inequality holds, and in particular if we have α-fractional log-concavity; we leave the details to Section 4.

Next, we discuss the techniques underlying our results in Section 1.3. A classic approach to the analysis of Markov chain mixing times consists of establishing functional inequalities, where roughly speaking, one shows that a measure of distance to the stationary measure μ multiplicatively contracts at every step. Two popular measures of distance to stationarity for a distribution ν are the χ^2-divergence, a.k.a. the variance of ν's density w.r.t. μ:

$$\mathbb{E}_\mu \left[\left(\frac{d\nu}{d\mu} - 1 \right)^2 \right],$$

and the relative entropy, a.k.a. the Kullback-Leibler divergence:

$$\mathbb{E}_\mu \left[\frac{d\nu}{d\mu} \cdot \log \frac{d\nu}{d\mu} \right].$$

Contraction of these "divergences" are related to Poincare and modified log-Sobolev inequalities respectively [see, e.g., BT06]. Contraction of variance is often easier to establish, because of its relation to the spectral gap of the Markov chain which enables a host of techniques for spectral analysis, but often it leads to a suboptimal (with a polynomial factor loss) bound on the mixing time; in contrast, modified log-Sobolev inequalities are notoriously difficult to establish, especially since there is no equivalent spectral connection, but they can lead to optimal mixing time bounds.

It is well-known that entropy contraction is strictly stronger than variance contraction [BT06]. Moreover, for distributions ν that are infinitesimally close to μ, entropy contraction and variance contraction become *equivalent*. Roughly speaking, this is because the functional $x \mapsto x \log x$ can be approximated by its quadratic Taylor expansion near $x = 1$, with the second degree term giving us the variance.

Restricted modified log-Sobolev inequalities. Motivated by the observation that entropy contraction and variance contraction are equivalent in infinitesmially small neighborhoods of the stationary distribution, we propose studying an intermediate form of functional inequality that we call a *restricted modified log-Sobolev inequality*. Roughly speaking, this is an inequality which guarantees entropy contraction in one step of the Markov chain for a *restricted class* of distribtuions ν. Intuitively, one should think of this as entropy contraction in a (sufficiently) *large neighborhood* of the stationary distribution. Our work shows that, in well-studied settings, restricted modified log-Sobolev inequalities can be considerably easier to establish than (full) modified log-Sobolev inequalities, while at the same time, yielding essentially the same consequences for mixing times and concentration of measure.

Restricted entropic independence. In order to establish restricted modified log-Sobolev inequalities we use a generalization of the notion of entropic independence (Definition 2). In Theorem 4, we show that spectral independence [ALO20], a form of variance contraction, for not just the distribution μ, but rather all external fields applied to μ, automatically entails entropic independence, a form of entropy contraction. The main barrier in applying this framework to the hardcore model, is that arbitrary external fields can easily take us outside the uniqueness regime where there is no hope of mixing, let alone spectral independence;

this is because an external field can change the parameter λ (the fugacity) to an arbitrarily large positive number. Nevertheless, we employ the fact that a *restricted* class of external fields keep the distribution in the spectral independence regime [Che+21a], and generalize the entropic independence machinery to show entropy contraction for a *restricted* class of distributions ν, which includes all of the distributions necessary for analyzing the mixing time of Markov chains and concentration of Lipschitz functions.

Boosting contraction results using field dynamics. We follow the footsteps of the prior work of Chen, Feng, Yin, and Zhang [Che+21a] who invented a new Markov chain called field dynamics, and showed its utility in establishing a spectral gap, both for the field dynamics itself, and by a comparison argument, for the Glauber dynamics. Field dynamics allows one to combine a loose bound on variance contraction near the uniqueness threshold together with an optimal bound for variance contraction far away from the threshold, to get a boosted optimal bound on variance contraction near the threshold. Our arguments follow the same high-level plan but with variance replaced with entropy. That is, we establish restricted modified log-Sobolev inequalities for the field dynamics first, and use optimal entropy contraction inequalities far away from the uniqueness threshold, to get a boosted optimal entropy contraction near the threshold. We then use comparison arguments to translate the results to a variant of Glauber dynamics.

A challenging part of using restricted modified log-Sobolev inequalities to establish mixing times is that a priori there is no reason that the evolution of the Markov chain will keep the distribution in the restricted class where we have entropy contraction, even if we initially start from a distribution within this class. We show that in the case of tree-unique hardcore and Ising models, simple modifications of the well-studied Glauber dynamics Markov chain and the field dynamics guarantee that the distribution at time t never escapes the restricted class of distributions.

Balanced Glauber dynamics, and field dynamics interleaved with systematic scans. As noted above, in our analysis of the hardcore model, we consider a variant of Glauber dynamics. In this variation, the update site is chosen in a slightly non-uniform fashion, effectively introducing a small number of additional "balancing" updates into the usual Glauber chain. Similarly, for another Markov chain called field dynamics that was introduced by Chen, Feng, Yin, and Zhang [Che+21a], we sometimes add an additional interleaving systematic scan step to keep the distribution within the restricted region of entropy contraction.

The introduction of these additional steps is very analogous to the use of *projections* in optimization algorithms such as projected gradient descent. In our case, these steps serve as projection operators in the following sense: they guarantee that the density of the resulting distribution lies in a restricted class of measures where we have contraction of entropy, while ensuring that the projection itself does not increase the relative entropy. The projection step enables us to show that the Glauber/field dynamics step makes a large amount of progress. In the optimization literature, such projection steps are sometimes crucial: the Iterative Hard Thresholding algorithm [BD09] alternates between a projection onto the set of sparse vectors and a gradient step on the squared loss, where the sparsity generated by the projection step is needed to argue that the gradient step makes progress (enabling appeal to the "Restricted Isometry Property"). Somewhat similarly, the Nash-Moser iteration (see, e.g., [Sec16]) combines the Newton step with a step which improves regularity.

We leave it as an interesting open question to investigate whether for the hardcore model, the balancing steps added to Glauber dynamics are actually needed. Stated differently, does vanilla Glauber dynamics (potentially started from a judicious choice of starting point) automatically remain in the C-bounded region of entropy contraction?

Concentration inequalities. MLSIs have other applications beyond mixing time of Markov chains; for example, they can be used to establish concentration inequalities using a technique known as the Herbst argument [see, e.g., Goe04]. We show that for the Ising model and the hardcore model in the uniqueness region, restricted modified log-Sobolev inequalities are enough to establish the same optimal concentration inequalities (as would be obtained by conjectured modified log-Sobolev inequalities), by demonstrating that

the Herbst argument essentially only needs entropy contraction for functionals within the "good restricted class" of C-bounded measures.

Sublinear time sampling algorithms. Our results improve the mixing time bounds for the high-degree regime of the Ising and hardcore models. One concern might be that mixing time could be a misleading indicator of algorithmic tractability; after all, it is easy to construct Markov chains that mix in one step, but whose steps take exponential time to implement. This concern is moot for Glauber dynamics in bounded-degree graphical models, as the steps of Glauber dynamics can be easily implemented in constant time. We show that this concern is moot even for the high-degree regime, by introducing new tricks to implement Glauber dynamics updates of the tree-unique Ising and hardcore models in amortized $O(1)$ time per step, improving on the naïve implementation which takes $O(\Delta)$ time per update. For the Ising model, we assume the ability to sample uniformly random neighbors of any desired node in the graph, and show that a trick based on Bernoulli factories can achieve the desired $O(1)$ update time. As far as we know, this trick has not been studied before, and it might be of independent interest.

1.5 Acknowledgment

Nima Anari and Thuy-Duong Vuong are supported by NSF CAREER award CCF-2045354, a Sloan Research Fellowship, and a Google Faculty Research Award. Frederic Koehler was supported in part by E. Mossel's Vannevar Bush Fellowship ONR-N00014-20-1-2826. Huy Tuan Pham is supported by a Two Sigma Fellowship.

2 Organization

The remainder of this extended abstract is organized as follows. In Section 4, we prove Theorem 4. The proof of this result relies on several preliminary notions, which we record in Section 3. The proofs of the results in Sections 1.1 and 1.2 may be found in the full version [Ana+21a], and the proofs of the results in Section 1.3 may be found in the full version [Ana+21b].

3 Preliminaries

3.1 Markov Chains and Functional Inequalities

Let μ and ν be probability measures on a finite set Ω. The Kullback-Liebler divergence (or relative entropy) between ν and μ is given by

$$\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) = \sum_{x \in \Omega} \nu(x) \log\left(\frac{\nu(x)}{\mu(x)}\right),$$

with the convention that this is ∞ if ν is not absolutely continuous with respect to μ. By Jensen's inequality, $\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \geq 0$ for any probability measures μ, ν. The total variation distance between μ and ν is given by

$$d_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

A Markov chain on Ω is specified by a row-stochastic non-negative transition matrix $P \in \mathbb{R}^{\Omega \times \Omega}$. We refer the reader to [LP17] for a detailed introduction to the analysis of Markov chains. As is common, we will view probability distributions on Ω as row vectors. Recall that a transition matrix P is said to be reversible with respect to a distribution μ if for all $x, y \in \Omega$, $\mu(x)P(x,y) = \mu(y)P(y,x)$. In this case, it follows immediately that μ is a stationary distribution for P i.e. $\mu P = \mu$. If P is further assumed to be ergodic, then μ is its unique stationary distribution, and for any probability distribution ν on Ω, $d_{\mathrm{TV}}(\nu P^t, \mu) \to 0$ as $t \to \infty$. The goal of this paper is to investigate the rate of this convergence.

Definition 22. Let P be an ergodic Markov chain on a finite state space Ω and let μ denote its (unique) stationary distribution. For any probability distribution ν on Ω and $\epsilon \in (0,1)$, we define

$$t_{\text{mix}}(P, \nu, \epsilon) = \min\{t \geq 0 \mid d_{\text{TV}}(\nu P^t, \mu) \leq \epsilon\},$$

and

$$t_{\text{mix}}(P, \epsilon) = \max\left\{\min\left\{t \geq 0 \mid d_{\text{TV}}(\mathbb{1}_x P^t, \mu) \leq \epsilon\right\} \mid x \in \Omega\right\},$$

where $\mathbb{1}_x$ is the point mass supported at x.

We will drop P and ν if they are clear from context. Moreover, if we do not specify ϵ, then it is set to $1/4$. This is because the growth of $t_{\text{mix}}(P, \epsilon)$ is at most logarithmic in $1/\epsilon$ (cf. [LP17]).

The modified log-Sobolev constant of a Markov chain, defined next, provides control on its mixing time.

Definition 23. Let P denote the transition matrix of an ergodic, reversible Markov chain on Ω with stationary distribution μ.

- The Dirichlet form of P is defined for $f, g : \Omega \to \mathbb{R}$ by

$$\mathcal{E}_P(f, g) = \langle f, (I - P)g\rangle_\mu = \langle (I - P)f, g\rangle_\mu.$$

- The modified log-Sobolev constant of P is defined to be

$$\rho_0(P) = \inf\left\{\frac{\mathcal{E}_P(f, \log f)}{2 \cdot \text{Ent}_\mu[f]} \;\middle|\; f \colon \Omega \to \mathbb{R}_{\geq 0}, \text{Ent}_\mu[f] \neq 0\right\},$$

  where

$$\text{Ent}_\mu[f] = \mathbb{E}_\mu[f \log f] - \mathbb{E}_\mu[f] \log \mathbb{E}_\mu[f].$$

  Note that, by rescaling, the infimum may be restricted to functions $f \colon \Omega \to \mathbb{R}_{\geq 0}$ satisfying $\text{Ent}_\mu[f] \neq 0$ and $\mathbb{E}_\mu[f] = 1$.

The relationship between the modified log-Sobolev constant and mixing times is captured by the following well-known lemma.

Lemma 24 ([see, e.g., BT06])**.** *Let P denote the transition matrix of an ergodic, reversible Markov chain on Ω with stationary distribution μ and let $\rho_0(P)$ denote its modified log-Sobolev constant. Then, for any probability distribution ν on Ω and for any $\epsilon \in (0,1)$*

$$t_{\text{mix}}(P, \nu, \epsilon) \leq \left\lceil \rho_0(P)^{-1} \cdot \left(\log\log \max\left\{\left(\frac{\nu(x)}{\mu(x)}\right) \;\middle|\; x \in \Omega\right\} + \log\left(\frac{1}{2\epsilon^2}\right)\right)\right\rceil.$$

In particular,

$$t_{\text{mix}}(P, \epsilon) \leq \left\lceil \rho_0(P)^{-1} \cdot \left(\log\log\left(\frac{1}{\min\{\mu(x) \mid x \in \Omega\}}\right) + \log\left(\frac{1}{2\epsilon^2}\right)\right)\right\rceil.$$

The next lemma, which shows that contraction of relative entropy under P implies a modified log-Sobolev inequality, is standard.

Lemma 25. *Let μ be a probability measure on the finite set Ω. Let P denote the transition matrix of an ergodic, reversible Markov chain on Ω with stationary distribution μ. Suppose there exists some $\alpha \in (0,1]$ such that for all probability measures ν on Ω which are absolutely continuous with respect to μ, we have*

$$\mathcal{D}_{\text{KL}}(\nu P \parallel \mu P) \leq (1 - \alpha)\, \mathcal{D}_{\text{KL}}(\nu \parallel \mu).$$

Then,

$$\rho_0(P) \geq 2 \cdot \alpha.$$

3.2 Down-Up Random Walks

Let $0 \le k \le n$ and let $\mu\colon \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ be a non-negative function on the size-k subsets of $[n]$. Note that μ is naturally associated to a probability distribution on $\binom{[n]}{k}$. We will find it useful to view the Ising model with n spins as a distribution over the size-n subsets of $[2n]$. For a set $\Omega = \{i_1, \dots, i_n\}$, we define the set $\bar{\Omega} = \{\bar{i_1}, \dots, \bar{i_n}\}$, which is disjoint from Ω, and each of whose elements is naturally paired with an element of Ω.

Definition 26. For $\sigma \in \{\pm 1\}^{\Omega}$, let $\sigma^{\hom} \in \binom{\Omega \cup \bar{\Omega}}{n}$ be the set $\{i \in \Omega \mid \sigma_i = 1\} \cup \left\{\bar{i} \in \bar{\Omega} \;\middle|\; \sigma_i = -1\right\}$. For a distribution μ on $\{\pm 1\}^{\Omega}$ with $|\Omega| = n$, let the homogenization of μ, denoted by $\mu^{\hom}$, be the distribution on $\left\{\sigma^{\hom} \;\middle|\; \sigma \in \{\pm 1\}^{\Omega}\right\}$ defined by $\mu^{\hom}(\sigma^{\hom}) \propto \mu(\sigma)$.

We will also find it useful to interpret the Glauber dynamics as the $n \leftrightarrow (n-1)$ down-up walk on $\binom{\Omega \cup \bar{\Omega}}{n}$. Recall that the down-up walk is given by the composition of two row-stochastic operators, known as the down and up operators.

Definition 27 (Down operator)**.** For a ground set Ω, and $|\Omega| \ge k \ge \ell$, define the down operator $D_{k \to \ell} \in \mathbb{R}^{\binom{\Omega}{k} \times \binom{\Omega}{\ell}}$ as

$$D_{k\to\ell}(S,T) = \begin{cases} \frac{1}{\binom{k}{\ell}} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $D_{k\to\ell} D_{\ell \to m} = D_{k \to m}$.

Definition 28 (Up operator)**.** For a ground set Ω, $|\Omega| \ge k \ge \ell$, and density $\mu : \binom{\Omega}{k} \to \mathbb{R}_{\ge 0}$, define the up operator $U_{\ell \to k} \in \mathbb{R}^{\binom{\Omega}{\ell} \times \binom{\Omega}{k}}$ as

$$U_{\ell\to k}(T,S) = \begin{cases} \frac{\mu(S)}{\sum_{S' \supseteq T} \mu(S')} & \text{if } T \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

If we define $\mu_k = \mu$ and more generally let μ_ℓ be $\mu_k D_{k\to\ell}$, then the down and up operators satisfy

$$\mu_k(S) D_{k\to\ell}(S,T) = \mu_\ell(T) U_{\ell\to k}(T,S).$$

This property ensures that the composition of the down and up operators have the appropriate μ as a stationary distribution, are reversible, and have nonnegative real eigenvalues.

Proposition 29 ([see, e.g., KO18; AL20; ALO20])**.** *The operators $D_{k\to\ell}U_{\ell\to k}$ and $U_{\ell\to k}D_{k\to\ell}$ both define Markov chains that are time-reversible and have nonnegative eigenvalues. Moreover μ_k and μ_ℓ are respectively their stationary distributions.*

Definition 30 (Down-up walk)**.** For a ground set Ω, $|\Omega| \ge k \ge \ell$, and density $\mu : \binom{\Omega}{k} \to \mathbb{R}_{\ge 0}$, the $k \leftrightarrow \ell$ down-up walk is defined by the row-stochastic matrix $U_{\ell\to k} D_{k\to\ell}$.

3.3 Polynomials

Definition 31. The multivariate generating polynomial g_μ associated to a density $\mu\colon \binom{[n]}{k} \to \mathbb{R}_{\ge 0}$ is the element of $\mathbb{R}[z_1, \dots, z_n]$ given by

$$g_\mu(z_1, \dots, z_n) := \sum_S \mu(S) \prod_{i \in S} z_i = \sum_S \mu(S) z^S,$$

Here we have used the standard notation that for $S \subseteq [n]$, $z^S = \prod_{i \in S} z_i$.

In [Ali+21], the notion of fractional log-concavity of the multivariate generating polynomial was developed. We will need a slight generalization of this notion.

Definition 32 (Non-uniform fractional log-concavity)**.** Consider a homogeneous distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ and let $g_\mu(z_1, \ldots, z_n)$ be its multivariate generating polynomial. For $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in [0,1]^n$, we say that μ is $\vec{\alpha}$-fractionally log-concave (abbreviated as $\vec{a}$-FLC) if $\log g_\mu(z_1^{\alpha_1}, \ldots, z_n^{\alpha_n})$ is concave, viewed as a function over $\mathbb{R}^n_{\geq 0}$.

Remark 33. If the distribution μ on $\binom{[n]}{k}$ is $\vec{\alpha}$-FLC, then the same is true for the conditional distributions μ_T for all $T \subseteq \binom{[n]}{\leq k}$. Here, μ_T is the distribution on $\binom{[n]\setminus T}{k-|T|}$ with $\mu_T(S) \propto \mu(T \cup S)$. This is because

$$g_{\mu_T} \propto \lim_{\lambda \to \infty} \lambda^{-|T|} \cdot g_\mu(\{\lambda z_i\}_{i \in T}, \{z_i\}_{i \notin T})$$

and operations of scaling the variables or the polynomial and taking limits preserve $\vec{\alpha}$-FLC.

We will use the following characterization of log-concavity for homogeneous functions. Recall that a function $f : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be d-homogeneous if $f(cx) = c^d f(x)$ for all $c > 0$.

Lemma 34 (Folklore)**.** *Let $\mathcal{C} \subseteq \mathbb{R}^n_{\geq 0}$ denote a convex cone. For a d-homogeneous function $f : \mathcal{C} \to \mathbb{R}_{\geq 0}$ the following are all equivalent:*

1. *f is quasi-concave.*
2. *f is log-concave.*
3. *f is d-th-root-concave, i.e., $f^{1/d}$ is concave.*

We will also need the following characterization of the solution of the minimum relative entropy problem with prescribed marginals which can be obtained by writing down the dual program and verifying Slater's condition.

Lemma 35 ([see, e.g., SV14, Appendix B])**.** *Consider a homogeneous distribution $\mu : \binom{[n]}{k} \to \mathbb{R}_{\geq 0}$ and let $g_\mu(z_1, \ldots, z_n)$ be its multivariate generating polynomial. Then, for any $q \in \mathbb{R}^n_{\geq 0}$ with $\sum_{i=q}^n q_i = 1$, we have*

$$\inf \{\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) \mid \nu D_{k \to 1} = q\} = -\log \left(\inf_{z_1, \ldots, z_n > 0} \frac{g_\mu(z_1, \ldots, z_n)}{z_1^{kq_1} \cdots z_n^{kq_n}} \right).$$

4 Fractional log-concavity and entropic independence

We now prove Theorem 4. Given this result, Theorem 5 follows from a version of the local-to-global argument, as discussed earlier; the proof is given in the Appendix in the full version [Ana+21a].

Proof of Theorem 4. Let $\alpha \in (0, 1]$ and let

$$\dagger : \forall (z_1, \ldots, z_n) \in \mathbb{R}^n_{\geq 0},\ g_\mu(z_1, \ldots, z_n) \leq \left(\sum_{i=1}^n p_i z_i^{1/\alpha} \right)^{\alpha k}$$

be the condition appearing in the statement of the theorem.

We first show that α-FLC implies ($\dagger$). Let μ be an arbitrary α-FLC distribution on $\binom{[n]}{k}$ and let $p := \mu D_{k \to 1} \in \mathbb{R}^n_{\geq 0}$. Note that

$$\partial_i g_\mu(1, \ldots, 1) = \frac{\sum_{S \ni i} \mu(S)}{\sum_S \mu(S)} = k p_i.$$

Since $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is αk-homogeneous and log-concave as a function of $z_1, \ldots, z_n$ over $\mathbb{R}^n_{\geq 0}$, it follows that

$$f(z_1, \ldots, z_n) := g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{1/\alpha k}$$

is 1-homogeneous and quasi-concave, hence, concave (Lemma 34). In particular,

$$\forall z_1,\dots,z_n>0: f(z_1,\dots,z_n)\le f(1,\dots,1)+\sum_{i=1}^n \partial_i f(1,\dots,1)(z_i-1).$$

By 1-homogeneity of f, $\sum_{i=1}^n \partial_i f(1,\dots,1)=f(1,\dots,1)$, so that

$$\forall z_1,\dots,z_n>0: f(z_1,\dots,z_n)\le \sum_{i=1}^n \partial_i f(1,\dots,1) z_i.$$

Moreover, since

$$\partial_i f(1,\dots,1)=\left(\alpha\cdot\partial_i g_\mu(1,\dots,1)\right)\cdot\left(\frac{1}{\alpha k}\cdot g_\mu(1,\dots,1)^{1/\alpha k-1}\right)=p_i,$$

we get that

$$\forall z_1,\dots,z_n>0: f(z_1,\dots,z_n)\le \sum_i p_i z_i.$$

Rewriting this in terms of g_μ yields

$$\forall z_1,\dots,z_n>0: g_\mu(z_1,\dots,z_n)\le \left(\sum_i p_i z_i^{1/\alpha}\right)^{\alpha k},$$

and now, (†) follows by continuity.

Next, we show that (†) implies $(1/\alpha)$-entropic independence. Let ν be an arbitrary distribution on $\binom{[n]}{k}$ and let $q:=\nu D_{k\to 1}$, so that $q\in\mathbb{R}^n_{\ge 0}$ with $\sum_{i=1}^n q_i=1$. We have from Lemma 35 that

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\nu\parallel\mu)&\ge \inf\left\{\mathcal{D}_{\mathrm{KL}}(\nu\parallel\mu)\mid \nu D_{k\to 1}=q\right\}\\&=-\log\left(\inf_{z_1,\dots,z_n>0}\frac{g_\mu(z_1,\dots,z_n)}{z_1^{kq_1}\cdots z_n^{kq_n}}\right).\end{aligned}$$

By (†),

$$\inf_{z_1,\dots,z_n>0}\frac{g_\mu(z_1,\dots,z_n)}{z_1^{kq_1}\cdots z_n^{kq_n}}\le \inf_{z_1,\dots,z_n>0}\frac{\left(\sum_i p_i z_i^{1/\alpha}\right)^{\alpha k}}{z_1^{kq_1}\cdots z_n^{kq_n}}.$$

Plugging in $z_i=(q_i/p_i)^\alpha$, we obtain

$$\inf_{z_1,\dots,z_n>0}\frac{g_\mu(z_1,\dots,z_n)}{z_1^{kq_1}\cdots z_n^{kq_n}}\le\prod_{i=1}^n (p_i/q_i)^{\alpha k q_i}.$$

Taking log and negating gives

$$\mathcal{D}_{\mathrm{KL}}(\nu\parallel\mu)\ge \alpha k\sum_i q_i\log(q_i/p_i)=\alpha k\cdot\mathcal{D}_{\mathrm{KL}}(\nu D_{k\to 1}\parallel\mu D_{k\to 1}).$$

Since ν is arbitrary, we obtain $(1/\alpha)$-entropic independence.

Now, we show that $(1/\alpha)$-entropic independence implies (†). By induction on the lower-dimensional faces of the positive orthant and homogeneity, it suffices to show that

$$\ell(z_1,\dots,z_n):=\left(\sum_i p_i z_i^{1/\alpha}\right)^{\alpha k}\ge 1\quad\forall z=(z_1,\dots,z_n)\in C,$$

where

$$C=\{(z_1,\dots,z_n): g_\mu(z_1,\dots,z_n)=1\wedge\nabla(\ell/g_\mu)(z_1,\dots,z_n)=0\}.$$

Let $z^* = (z_1^*, \ldots, z_n^*) \in C$. Then, $\nu = z^* * \mu$ is a distribution on $\binom{[n]}{k}$ with $q = (q_1, \ldots, q_n) := \nu D_{k \to 1} \propto (p_1 (z_1^*)^{1/\alpha}, \ldots, p_n (z_n^*)^{1/\alpha})$. Now, examining the first order condition shows that

$$\begin{aligned}\mathcal{D}_{\mathrm{KL}}(\nu \parallel \mu) &= -\log\left(\inf_{z_1,\ldots,z_n>0} \frac{g_\mu(z_1,\ldots,z_n)}{z_1^{kq_1}\ldots z_n^{kq_n}}\right) \\ &= -\log\left(\frac{1}{(z_1^*)^{kq_1}\ldots(z_n^*)^{kq_n}}\right),\end{aligned}$$

so that by $(1/\alpha)$-entropic independence,

$$\frac{1}{(z_1^*)^{kq_1}\ldots(z_n^*)^{kq_n}} \le \prod_{i=1}^n (p_i/q_i)^{\alpha k q_i} = \frac{\ell(z_1^*,\ldots,z_n^*)}{(z_1^*)^{kq_1}\ldots(z_n^*)^{kq_n}},$$

from which we get that $\ell(z_1^*, \ldots, z_n^*) \ge 1$, as desired.

Finally, we establish the equivalence between entropic independence under arbitrary external fields and fractional log-concavity. In one direction, we note that α-fractional log-concavity of μ immediately implies α-fractional log-concavity of $\lambda * \mu$ for any $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{>0}$, which as we have just seen, implies $(1/\alpha)$-entropic independence of $\lambda * \mu$. In the other direction, suppose that $\lambda * \mu$ is $(1/\alpha)$-entropic independent for all $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{>0}$. Then, using (†) for all $\lambda * \mu$, we see that

$$\forall z_1, \ldots, z_n > 0 : g_\mu(z_1^\alpha, \ldots, z_n^\alpha)^{1/\alpha k} = \inf_{\lambda \in \mathbb{R}^n_{>0}} \sum_i p(\lambda)_i z_i.$$

Since a pointwise infimum of linear functions is concave, it follows that $g_\mu(z_1^\alpha, \ldots, z_n^\alpha)$ is αk-root-concave, and hence, log-concave (Lemma 34). This completes the proof. □

References

[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.

[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: *arXiv preprint arXiv:2102.02708* (2021).

[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.

[Ana+19] 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+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: (2021). DOI: 10.48550/ARXIV.2106.04105.

[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: *CoRR* abs/2111.03247 (2021). arXiv: 2111.03247.

[AV21] Nima Anari and Thuy-Duong Vuong. "From Sampling to Optimization on Discrete Domains with Applications to Determinant Maximization". In: *arXiv preprint arXiv:2102.05347* (2021).

[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.

[BD09] Thomas Blumensath and Mike E Davies. "Iterative hard thresholding for compressed sensing". In: *Applied and computational harmonic analysis* 27.3 (2009), pp. 265–274.

[BG99] Sergej G Bobkov and Friedrich Götze. "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities". In: *Journal of Functional Analysis* 163.1 (1999), pp. 1–28.

[BH19] Petter Brändén and June Huh. "Lorentzian polynomials". In: *arXiv preprint arXiv:1902.03719* (2019).

[Bla+21] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Stefankovic, and Eric Vigoda. "On mixing of Markov chains: Coupling, spectral independence, and entropy factorization". In: *arXiv preprint arXiv:2103.07459* (2021).

[BRZ95] Pavel M Bleher, Jean Ruiz, and Valentin A Zagrebnov. "On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice". In: *Journal of Statistical Physics* 79.1 (1995), pp. 473–482.

[BT06] Sergey G Bobkov and Prasad Tetali. "Modified logarithmic Sobolev inequalities in discrete settings". In: *Journal of Theoretical Probability* 19.2 (2006), pp. 289–336.

[CGM20] Mary Cryan, Heng Guo, and Giorgos Mousa. *Modified log-Sobolev inequalities for strongly log-concave distributions*. 2020. arXiv: `1903.06081 [math.PR]`.

[Cha+86] JT Chayes, L Chayes, James P Sethna, and DJ Thouless. "A mean field spin glass with short-range interactions". In: *Communications in Mathematical Physics* 106.1 (1986), pp. 41–89.

[Che+21a] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. *Rapid mixing of Glauber dynamics via spectral independence for all degrees*. 2021. arXiv: `2105.15005 [cs.DS]`.

[Che+21b] Zongchen Chen, Andreas Galanis, Daniel tefankovi, and Eric Vigoda. "Rapid mixing for colorings via spectral independence". In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 1548–1557.

[CLV20] Zongchen Chen, Kuikui Liu, and Eric Vigoda. "Optimal Mixing of Glauber Dynamics: Entropy Factorization via High-Dimensional Expansion". In: *CoRR* abs/2011.02075 (2020). arXiv: `2011.02075`.

[DFJ02] Martin Dyer, Alan Frieze, and Mark Jerrum. "On counting independent sets in sparse graphs". In: *SIAM Journal on Computing* 31.5 (2002), pp. 1527–1541.

[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.

[EKZ20] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. "A Spectral Condition for Spectral Gap: Fast Mixing in High-Temperature Ising Models". In: *arXiv preprint arXiv:2007.08200* (2020).

[Eva+00] William Evans, Claire Kenyon, Yuval Peres, and Leonard J Schulman. "Broadcasting on trees and the Ising model". In: *Annals of Applied Probability* (2000), pp. 410–433.

[Fen+21] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. "Rapid mixing from spectral independence beyond the Boolean domain". In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 1558–1577.

[Gar+19] Mike Gartrell, Victor-Emmanuel Brunel, Elvis Dohmatob, and Syrine Krichene. "Learning nonsymmetric determinantal point processes". In: *arXiv preprint arXiv:1905.12962* (2019).

[Goe04] Sharad Goel. "Modified logarithmic Sobolev inequalities for some models of random walk". In: *Stochastic processes and their applications* 114.1 (2004), pp. 51–79.

[GV16] Andreas Galanis, Daniel tefankovi, 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.

[Hay06] Thomas P Hayes. "A simple condition implying rapid mixing of single-site dynamics on spin systems". In: *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*. IEEE. 2006, pp. 39–46.

[Iof96] Dmitry Ioffe. "On the extremality of the disordered state for the Ising model on the Bethe lattice". In: *Letters in Mathematical Physics* 37.2 (1996), pp. 137–143.

[JPV21] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. "Spectral independence, coupling with the stationary distribution, and the spectral gap of the Glauber dynamics". In: *arXiv preprint arXiv:2105.01201* (2021).

[KM16] Tali Kaufman and David Mass. "High dimensional random walks and colorful expansion". In: *arXiv preprint arXiv:1604.02947* (2016).

[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.

[Lau96] Steffen L Lauritzen. *Graphical models*. Vol. 17. Clarendon Press, 1996.

[Liu21] Kuikui Liu. "From Coupling to Spectral Independence and Blackbox Comparison with the Down-Up Walk". In: *CoRR* abs/2103.11609 (2021). arXiv: 2103.11609.

[LLP17] Eyal Lubetzky, Alex Lubotzky, and Ori Parzanchevski. "Random walks on Ramanujan complexes and digraphs". In: *arXiv preprint arXiv:1702.05452* (2017).

[LP17] David A Levin and Yuval Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.

[Mar19] Katalin Marton. "Logarithmic Sobolev inequalities in discrete product spaces". In: *Combinatorics, Probability and Computing* 28.6 (2019), pp. 919–935.

[Mar99] Fabio Martinelli. "Lectures on Glauber dynamics for discrete spin models". In: *Lectures on probability theory and statistics*. Springer, 1999, pp. 93–191.

[MM09] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[MOS13] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. "On reverse hypercontractivity". In: *Geometric and Functional Analysis* 23.3 (2013), pp. 1062–1097.

[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.

[MS13] Elchanan Mossel and Allan Sly. "Exact thresholds for Ising–Gibbs samplers on general graphs". In: *The Annals of Probability* 41.1 (2013), pp. 294–328.

[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.

[PP10] Robin Pemantle and Yuval Peres. "The critical Ising model on trees, concave recursions and nonlinear capacity". In: *The Annals of Probability* 38.1 (2010), pp. 184–206.

[Sec16] Paolo Secchi. "On the Nash-Moser iteration technique". In: *Recent developments of mathematical fluid mechanics*. Springer, 2016, pp. 443–457.

[Sly10] Allan Sly. "Computational transition at the uniqueness threshold". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE. 2010, pp. 287–296.

[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.

[SST14] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. "Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs". In: *Journal of Statistical Physics* 155.4 (2014), pp. 666–686.

[SV14] Mohit Singh and Nisheeth K Vishnoi. "Entropy, optimization and counting". In: *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. 2014, pp. 50–59.

[Tal10] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*. Vol. 54. Springer Science & Business Media, 2010.

[Van14] Ramon Van Handel. *Probability in high dimension*. Tech. rep. PRINCETON UNIV NJ, 2014.

[Var] Various. *Simons Institute Wiki (Fall 2020)*. https://wiki.simons.berkeley.edu/doku.php?id=hd20:glauber-dynamics-for-sk. Accessed: 2021-06-01.

[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.